Advanced Electrodynamics

Robert van Leeuwen

December 16, 2013

Preface

Contents

Chapter 1

Relativistic coordinate transformations

In this Chapter we will discuss which linear coordinate transformations are possible between reference frames moving uniformly with respect to each other. We will focus in particular on Galilean and Lorentz transformations and will address the way reference frames are constructed using Einstein synchronization of clocks. Some familiar physical consequences of Lorentz transformations are discussed and the important concept of proper time is introduced.

1.1 Galilean symmetry

Let us start out by specifying a coordinate system for measuring particles in space and time. To measure spatial position we choose three orthogonal coordinate axes. This assumes that any observer can physically construct a linear space with a standard Euclidean inner product (we will later relax this assumption). The position of objects with respect to the origin O can be determined by marking off distances along the axes with a unit measuring rod of length 1, i.e. on the x-axis the value x represents x times the length of our measuring rod. This procedure assigns to every point in space the triple (x, y, z). The distance L between two points (x, y, z) and $(x', y'z')$ is then given by the familiar Euclidean distance

$$L^2 = (x - x')^2 + (y - y')^2 + (z - z')^2$$

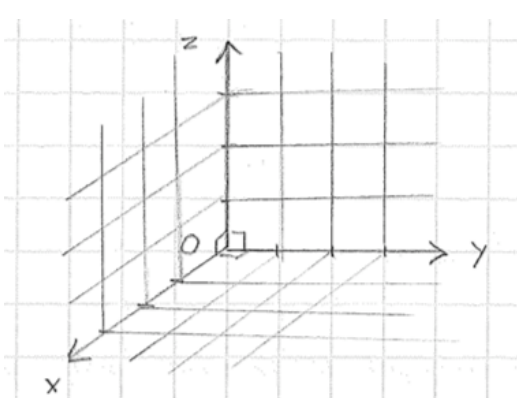

Figure 1.1: Euclidean coordinate system

Then we can assign to each such a point a time value t as well. The time at each point is measured by a local clock at that point which is synchronized with a clock in the origin O. We will later wonder in Section 1.3 how such a synchronization can be achieved, but it should already be kept in mind that this is an important conceptual issue since whether the laws of physics attain a simple form in our coordinates or not depends on our choice of clock synchronization. However, let us assume that we succeeded In this way an object at a certain position in the spatial coordinate frame is assigned a time coordinate as well. We denote such a space-time point by (t, x, y, z). Given the coordinate system we can talk about the motion of objects, where a motion is defined to be a one-dimensional curve in our coordinate system parametrized by t, i.e. change of position with time. Let us for simplicity assume that this object is a point particle. Then its motion is described by a trajectory

$$\mathbf{x}(t) = (x(t), y(t), z(t)).$$

The velocity $\mathbf{v}(t)$ of the particle is defined to be the time-derivative of its position, i.e.

$$\mathbf{v}(t) = \frac{d\mathbf{x}}{dt}(t) = (\frac{dx}{dt}(t), \frac{dy}{dt}(t), \frac{dz}{dt}(t))$$

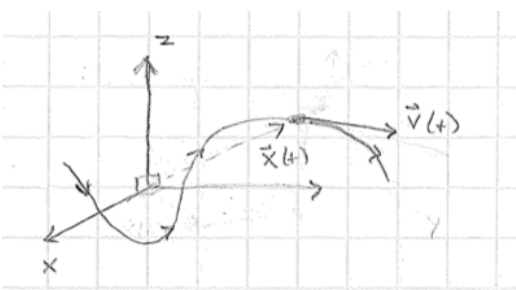

Figure 1.2: Velocity vector is tangent to the path of the particle

This vector is always tangent to the trajectory (see figure). Now that we can describe motion we can study physical processes, such as the motion of colliding billiard balls, the motion of planets in the Solar system etc.

There exists a special kind of reference frames, which are called *inertial frames*. Such frames are defined by the property that they are homogenous in space and time as well as isotropic. This means that any experiment carried out in them (such as letting two billiard balls collide) gives the same physical result independent of the position and orientation of the experimental apparatus as well of the time at which the experiment is carried out. It is an experimental fact that in such frames isolated objects, i.e. objects which do not interact with other objects, move at constant velocity, i.e $\mathbf{v}$ is time-independent and the motion represents a straight line (any simpler motion is hardly imaginable). Inertial frames have the property that any frame moving at constant velocity with respect to them is another inertial frame. The question which physical systems actually do form inertial frames in practice is addressed in Chapter 2. It is, however, easy to find non-inertial frames. For instance, an observer who is at rest with respect to a rotating disc will find that the motion of a billiard ball after it is released from an initial state at rest is dependent on the position on the disc from which it is released. This reference frame therefore violates the requirement of spatial homogeneity.

Let us assume that our reference frame is an inertial frame. Since in our frame we can describe motion we can, in particular, describe motion of other observers. Let us therefore consider two observes O_1 and O_2. Observers O_1 uses coordinates (t_1, x_1, y_1, z_1) and observer O_2 uses coordinates (t_2, x_2, y_2, z_2). The coordinate systems are chosen such that the systems coincide at time $t_1 = 0$ as measured by O_1, with their spatial coordinate axes parallel. Let us further assume that the system 2 moves with constant velocity along the x_1-axis of system 1 (see figure 1.3).

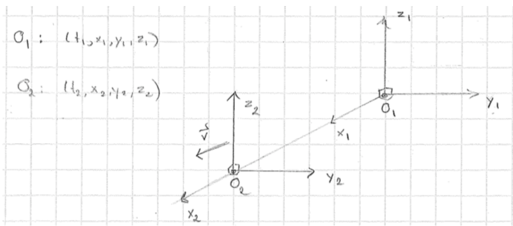

Figure 1.3: Relative motion of the coordinate frames

Constant velocity implies $d\mathbf{v}/dt = 0$ and we have

$$\mathbf{v}(t_1) = \frac{d\mathbf{x}}{dt_1} = (\frac{dx_1}{dt_1}, 0, 0) = (v, 0, 0) = \mathbf{v}$$

where v is a constant and $\mathbf{v}$ a constant vector. Therefore the position of observer O_2 (who is assumed to be in the origin of his coordinate system) as measured in system O_1 is given by

$$\mathbf{x}(t_1) = (vt_1, 0, 0)$$

or $x_1(t_1) = vt_1$, and the space-time coordinate of this point is given by $(t_1, vt_1, 0, 0)$. Observer O_2 has its own coordinate system in which time and position are measured. A space-time point in this system is denoted (t_2, x_2, y_2, z_2) and in particular the origin is given by $(t_2, 0, 0, 0)$. It now only remains to relate the time coordinates of both reference frames.
In Newtonian mechanics it is now assumed that there exists a universal clock that any observer at any spatial point can read instantaneously. Therefore any local clock at any position in the reference frame of any observer can be synchronized[1] with the universal clock displaying time t. If we do this then all relatively moving clocks display the same universal time which Newton denoted as *absolute time*. Let us assume that this concept is correct, then we can set $t_1 = t_2 = t$ for our two observers. Then we see (see figure) that if observer O_2 measures a particle at position $\mathbf{x}_2(t)$ then at the same time t observer O_1 measures a particle at position $\mathbf{x}_1(t) = \mathbf{x}_2(t) + \mathbf{v}\,t$ or equivalently

$$\mathbf{x}_2(t) = \mathbf{x}_1(t) - \mathbf{v}\,t.$$

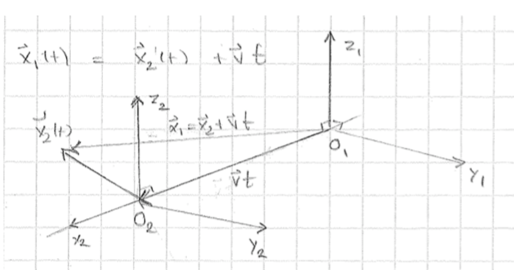

Figure 1.4: Relation between positions in different frames

[1]If one starts to think of ways of synchronizing a local clock with universal time in practice, one realizes that the existence of absolute time is not obvious at all.

More precisely, since we assume $t_1 = t_2$, the transformation from space-time coordinates (t_1, x_1, y_1, z_1) to (t_2, x_2, y_2, z_2) is given by

$$\begin{aligned} t_2 &= t_1 \\ x_2 &= x_1 - v\,t_1 \\ y_2 &= y_1 \\ z_2 &= z_1 \end{aligned} \tag{1.1}$$

This transformation is known as the *Galilean transformation*. The Galilean transformations form what is mathematically known as a *group*. To illustrate this we forget the y and z coordinates for a moment as they are unchanged anyway. The (t, x) coordinates in Eq.(1.1) then transform as follows

$$\begin{pmatrix} t_2 \\ x_2 \end{pmatrix} = \begin{pmatrix} t_1 \\ x_1 - v\,t_1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -v & 1 \end{pmatrix} \begin{pmatrix} t_1 \\ x_1 \end{pmatrix} = A(v) \begin{pmatrix} t_1 \\ x_1 \end{pmatrix} \tag{1.2}$$

where we defined the matrix

$$A(v) = \begin{pmatrix} 1 & 0 \\ -v & 1 \end{pmatrix} \tag{1.3}$$

This matrix has a number of obvious properties. First of all

$$A(0) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \mathbf{1} \tag{1.4}$$

is the unit matrix. This simply means that if the relative velocity of both systems is zero then then all coordinates are identical. Secondly

$$A^{-1}(v) = \begin{pmatrix} 1 & 0 \\ v & 1 \end{pmatrix} = A(-v) \tag{1.5}$$

This means that transforming back from coordinates (t_2, x_2) to (t_1, x_1) is the same transformation with minus the relative velocity. In other words, if O_2 moves with respect to O_1 with relative velocity v, then O_1 moves with respect to O_2 at relative velocity $-v$. A further property is

$$A(v_2)A(v_1) = \begin{pmatrix} 1 & 0 \\ -v_2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -v_1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -v_1 - v_2 & 1 \end{pmatrix} = A(v_1 + v_2) \tag{1.6}$$

This equation tells us that if system O_2 moves with relative velocity v_1 respect to O_1 and if system O_3 moves with relative velocity v_2 with respect to O_2 then system O_3 moves with relative velocity $v_3 = v_1 + v_2$ with respect to O_1. We therefore find the following formula for addition of velocities

$$v_3 = v_1 + v_2.$$

So, summarizing, we find that the Galilean transformations satisfy the properties

$$A(0) = \mathbf{1} \tag{1.7}$$
$$A^{-1}(v) = A(-v) \tag{1.8}$$
$$A(v_3) = A(v_2)A(v_1) \tag{1.9}$$

(where in Newtonian mechanics $v_3 = v_1 + v_2$). Properties (1.7)-(1.9) are properties of what mathematicians call a *group*. A group G is defined to be a set of elements with a multiplication $\cdot$ such that

1. If $g_1 \in G$ and $g_2 \in G$ then $g_1 \cdot g_2 \in G$ with the associativity property $(g_1 \cdot g_2) \cdot g_3 = g_1 \cdot (g_2 \cdot g_3)$
2. There is a unit element $\mathbf{1} \in G$ with the property $\mathbf{1} \cdot g = g \cdot \mathbf{1}$ for all $g \in G$
3. For any $g \in G$ there is an inverse $g^{-1} \in G$ with the property $g \cdot g^{-1} = g^{-1} \cdot g = \mathbf{1}$

It is clear that the set of Galilean transformations $G = \{A(v) | v \in \mathbb{R}\}$ satisfies all these axioms and therefore represents a group. There is a close relation between group structure and symmetry, since all operations that leave a symmetric object invariant (such as rotations and mirrorings) form a group. We may therefore refer to the set of Galilean transformations as the Galilean symmetry group.

1.2 Beyond Galilean symmetry

The coordinate transformation (1.1) was derived assuming that observers moving relative to each other measure the same time. We now want to relax this condition. We again consider simply motion in one spatial dimension and look for the most general transformation

$$\begin{pmatrix} t_2 \\ x_2 \end{pmatrix} = A(v) \begin{pmatrix} t_1 \\ x_2 \end{pmatrix}$$

that satisfies the properties (1.7), (1.8) and (1.9) and where v is the velocity of system O_2 with respect to O_1. Apart from these plausible properties we only make the assumption that space is isotropic, which means that there is no preferred direction in space. The matrix $A(v)$ is of the general form

$$A(v) = \begin{pmatrix} a(v) & b(v) \\ c(v) & d(v) \end{pmatrix}$$

where $a(v), b(v), c(v)$ and $d(v)$ are functions of the velocity that we need to determine. From

$$A(0) = \begin{pmatrix} a(0) & b(0) \\ c(0) & d(0) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

we find that $a(0) = d(0) = 1$ and $b(0) = c(0) = 0$. From

$$\mathbf{1} = A(v)A^{-1}(v) = A(v)A(-v) \tag{1.10}$$

it follows by taking the determinant that

$$1 = \det(A(v)A(-v)) = D(v)D(-v) \tag{1.11}$$

where we defined $D(v) = \det(A(v))$. Since we assumed that there is no preferred direction in space we must have that $D(v) = D(-v)$ (i.e. the scalar $D(v)$ does not distinguish between different directions of relative motion) and therefore

$$1 = D(v)D(-v) = D(v)^2$$

which implies that $D(v) = \pm 1$. However, since $D(0) = 1$ we must have that $D(v) = 1$. Property (1.8) tell us that

$$A(-v) = \begin{pmatrix} a(-v) & b(-v) \\ c(-v) & d(-v) \end{pmatrix} = \frac{1}{D(v)} \begin{pmatrix} d(v) & -b(v) \\ -c(v) & a(v) \end{pmatrix} = A^{-1}(v)$$

where $D(v) = a(v)d(v) - b(v)c(v)$ is the determinant of $A(v)$. Since we already showed that $D(v) = 1$ we find that

$$\begin{aligned} a(-v) &= d(v) && (1.12)\\ b(-v) &= -b(v) && (1.13)\\ c(-v) &= -c(v) && (1.14) \end{aligned}$$

We now use the fact that the origin $x_2 = 0$ of system O_2 has with respect to O_1 the position $x_1 = vt_1$, i.e.

$$\begin{pmatrix} t_2 \\ 0 \end{pmatrix} = A(v) \begin{pmatrix} t_1 \\ v\,t_1 \end{pmatrix}.$$

This yields the equation

$$0 = c(v)t_1 + d(v)\,v\,t_1$$

for all t_1. This implies that

$$c(v) = -vd(v) \tag{1.15}$$

which in turn, together with Eq.(1.14), implies that

$$vd(-v) = c(-v) = -c(v) = vd(v)$$

and therefore $d(v) = d(-v)$. From Eq.(1.12) we then see that $a(v) = d(v)$ and, together with Eq.(1.15), we obtain

$$A(v) = \begin{pmatrix} a(v) & b(v) \\ -va(v) & a(v) \end{pmatrix} \tag{1.16}$$

where $a(v) = a(-v)$ and $b(v) = -b(-v)$. Since $D(v) = 1$ we also know that

$$1 = D(v) = a^2(v) + v\,a(v)b(v)$$

and we therefore find that

$$b(v) = \frac{1 - a^2(v)}{v\,a(v)} \tag{1.17}$$

We have therefore reduced the problem to the determination of the unknown function $a(v)$. To find $a(v)$ we use the final condition (1.9) which tells us that

$$\begin{aligned} A(v_3) &= \begin{pmatrix} a(v_3) & b(v_3) \\ -v_3a(v_3) & a(v_3) \end{pmatrix} = A(v_2)A(v_1) \\ &= \begin{pmatrix} a(v_2) & b(v_2) \\ -v_2a(v_2) & a(v_2) \end{pmatrix} \begin{pmatrix} a(v_1) & b(v_1) \\ -v_1a(v_1) & a(v_1) \end{pmatrix} \\ &= \begin{pmatrix} a(v_1)a(v_2) - v_1a(v_1)b(v_2) & a(v_2)b(v_1) + a(v_1)b(v_2) \\ -(v_1 + v_2)a(v_1)a(v_2) & a(v_1)a(v_2) - v_2a(v_2)b(v_1) \end{pmatrix} \qquad (1.18) \end{aligned}$$

Since the diagonal elements of $A(v_3)$ are identical it follows that

$$-v_1\,a(v_1)b(v_2) = -v_2\,a(v_2)b(v_1)$$

or equivalently

$$\frac{b(v_1)}{v_1\,a(v_1)} = \frac{b(v_2)}{v_2\,a(v_2)}$$

which must be true for all possible choices of v_1 and v_2. This can only be true when

$$\frac{b(v)}{v\,a(v)} = K = \text{constant}.$$

In combination with Eq.(1.17) this immediately yields

$$\frac{1-a^2(v)}{v^2\,a^2(v)} = K \quad \Rightarrow \quad a(v) = \frac{\pm 1}{\sqrt{1+Kv^2}}. \tag{1.19}$$

However, since $a(0) = 1$ we must have

$$a(v) = \frac{1}{\sqrt{1+Kv^2}}$$

and consequently

$$b(v) = Kv\,a(v) = \frac{K\,v}{\sqrt{1+Kv^2}}.$$

The matrix of Eq.(1.16) therefore attains the form

$$A(v) = \frac{1}{\sqrt{1+Kv^2}} \begin{pmatrix} 1 & K\,v \\ -v & 1 \end{pmatrix}. \tag{1.20}$$

This is the final result of our derivation. It is clear that for $K = 0$ we recover the Galilean transformation (1.3). However, for $K \neq 0$ we obtain a new more general transformation in which both space and time coordinates are mixed:

$$\begin{aligned} t_2 &= \frac{1}{\sqrt{1+Kv^2}}(t_1 + Kv\,t_1) && (1.21) \\ x_2 &= \frac{1}{\sqrt{1+Kv^2}}(x_1 - v\,t_1). && (1.22) \end{aligned}$$

Nothing has been assumed about the sign of K but the transformation (1.20) is qualitatively different for $K > 0$ and $K < 0$. Let us investigate this a bit further. Since the constant K has the dimension $[1/\text{velocity}^2]$ we can write

$$K = \pm\frac{1}{c^2}$$

with $c > 0$ having the dimension of velocity. Let us first consider the case $K = 1/c^2$, i.e. we take K to be positive. Then Eqs.(1.21) and (1.22) become

$$\begin{aligned} t_2 &= \frac{1}{\sqrt{1+\frac{v^2}{c^2}}}(t_1 + \frac{v}{c^2}\,x_1) \\ x_2 &= \frac{1}{\sqrt{1+\frac{v^2}{c^2}}}(x_1 - v\,t_1). \end{aligned}$$

By defining $\tau_1 = ct_1$ and $\tau_2 = ct_2$ we can write these equations as

$$\begin{aligned} \begin{pmatrix} \tau_2 \\ x_2 \end{pmatrix} &= \frac{1}{\sqrt{1+\frac{v^2}{c^2}}} \begin{pmatrix} 1 & \frac{v}{c} \\ -\frac{v}{c} & 1 \end{pmatrix} \begin{pmatrix} \tau_1 \\ x_1 \end{pmatrix} \\ &= \begin{pmatrix} \cos(\phi) & \sin(\phi) \\ -\sin(\phi) & \cos(\phi) \end{pmatrix} \begin{pmatrix} \tau_1 \\ x_1 \end{pmatrix} = B(\phi) \begin{pmatrix} \tau_1 \\ x_1 \end{pmatrix} && (1.23) \end{aligned}$$

This equation defines the rotation matrix $B(\phi)$ dependent on the angle ϕ determined from $\tan(\phi) = v/c$ or equivalently

$$\phi = \arctan(\frac{v}{c}). \tag{1.24}$$

Redefining similarly $\tau_3 = ct_3$ we have from Eq.(1.9) that

$$\begin{aligned} B(\phi_3) &= B(\phi_2)B(\phi_1) = \begin{pmatrix} \cos(\phi_2) & \sin(\phi_2) \\ -\sin(\phi_2) & \cos(\phi_2) \end{pmatrix} \begin{pmatrix} \cos(\phi_1) & \sin(\phi_1) \\ -\sin(\phi_1) & \cos(\phi_1) \end{pmatrix} \\ &= \begin{pmatrix} \cos(\phi_1+\phi_2) & \sin(\phi_1+\phi_2) \\ -\sin(\phi_1+\phi_2) & \cos(\phi_1+\phi_2) \end{pmatrix} = B(\phi_1+\phi_2) \qquad (1.25) \end{aligned}$$

where we defined the angles $\phi_i = \arctan(v_i/c)$ for $i = 1, 2, 3$. We thus find that

$$\frac{v_3}{c} = \tan(\phi_1+\phi_2) = \frac{\tan(\phi_1)+\tan(\phi_2)}{1-\tan(\phi_1)\tan(\phi_2)} = \frac{\frac{v_1}{c}+\frac{v_2}{c}}{1-\frac{v_1 v_2}{c^2}}$$

or equivalently

$$v_3 = \frac{v_1+v_2}{1-\frac{v_1 v_2}{c^2}} \qquad (1.26)$$

This gives a rather unphysical formula for the addition of velocities. In particular, v_3 diverges for $v_1 v_2 \to c^2$ and changes sign for $v_1 v_2 > c^2$. In any case it is in disagreement with experiment and we consider the case $K > 0$ as physically unacceptable. Let us therefore turn to the case that $K = -1/c^2$ such that $K < 0$. Then Eqs.(1.21) and (1.22) become

$$\begin{aligned} t_2 &= \frac{1}{\sqrt{1-\frac{v^2}{c^2}}}(t_1 - \frac{v}{c^2}x_1) &(1.27) \\ x_2 &= \frac{1}{\sqrt{1-\frac{v^2}{c^2}}}(x_1 - v\,t_1). &(1.28) \end{aligned}$$

This transformation is known as the *Lorentz transformation*. We note that these equations make only sense when $|v| < c$ and hence assume this condition to be satisfied. By defining again $\tau_1 = ct_1$ and $\tau_2 = ct_2$ we can write

$$\begin{aligned} \begin{pmatrix} \tau_2 \\ x_2 \end{pmatrix} &= \frac{1}{\sqrt{1-\frac{v^2}{c^2}}} \begin{pmatrix} 1 & -\frac{v}{c} \\ -\frac{v}{c} & 1 \end{pmatrix} \begin{pmatrix} \tau_1 \\ x_1 \end{pmatrix} \\ &= \begin{pmatrix} \cosh(\phi) & -\sinh(\phi) \\ -\sinh(\phi) & \cosh(\phi) \end{pmatrix} \begin{pmatrix} \tau_1 \\ x_1 \end{pmatrix} = B(\phi) \begin{pmatrix} \tau_1 \\ x_1 \end{pmatrix} \qquad (1.29) \end{aligned}$$

where we defined the angle ϕ by

$$\tanh(\phi) = \frac{v}{c}.$$

From Eq.(1.9) we have again that

$$\begin{aligned} B(\phi_3) &= B(\phi_2)B(\phi_1) = \begin{pmatrix} \cosh(\phi_2) & -\sinh(\phi_2) \\ -\sinh(\phi_2) & \cosh(\phi_2) \end{pmatrix} \begin{pmatrix} \cosh(\phi_1) & -\sinh(\phi_1) \\ -\sinh(\phi_1) & \cosh(\phi_1) \end{pmatrix} \\ &= \begin{pmatrix} \cosh(\phi_1+\phi_2) & -\sinh(\phi_1+\phi_2) \\ -\sinh(\phi_1+\phi_2) & \cosh(\phi_1+\phi_2) \end{pmatrix} = B(\phi_1+\phi_2) \qquad (1.30) \end{aligned}$$

where we defined the angles $\phi_i = \mathrm{arctanh}(v_i/c)$ for $i = 1, 2, 3$. We thus find that

$$\frac{v_3}{c} = \tanh(\phi_1+\phi_2) = \frac{\tanh(\phi_1)+\tanh(\phi_2)}{1+\tanh(\phi_1)\tanh(\phi_2)} = \frac{\frac{v_1}{c}+\frac{v_2}{c}}{1+\frac{v_1 v_2}{c^2}}$$

We thus obtain the following addition theorem for velocities:

$$v_3 = \frac{v_1+v_2}{1+\frac{v_1 v_2}{c^2}}. \qquad (1.31)$$

An important difference with respect to the case of positive K is that the Lorentz transformations only make sense when $|v| < c$. Therefore the relative velocity of two physical systems can not be larger than c. What, however, happens if we add velocities according to Eq.(1.31) ? If we let $v_2 = \pm c$ we find

$$v_3 = \frac{v_1 \pm c}{1 \pm \frac{v_1}{c}} = \pm c.$$

So by adding velocities v_1, v_2 with $|v_1|, |v_2| < c$ we can never attain a value larger than c. Therefore c appears as a maximum attainable velocity. It turns out that nature is indeed well-described by a theory of this kind. In electromagnetism the velocity c is the velocity at which electromagnetic waves propagate in empty space. Since visible light is nothing but an electromagnetic wave (leaving out quantum theory for the moment) we may refer to c as the 'light speed'. However, this speed is not particular to light but also other fundamental forces than electromagnetism have force carriers that move at the same speed. The strong force is, for instance, mediated by a gluon field moving at 'light speed', and also gravity waves are moving at light speed. One thing that is immediately clear from Eqs. (1.27) and (1.28) is that light speed is universal for all observers. If for one observer a signal moves at speed c, i.e. $x_1 = ct_1$, then (1.27) and (1.28) tell us that

$$\begin{aligned} t_2 &= \frac{1}{\sqrt{1-\frac{v^2}{c^2}}}(t_1 - \frac{v}{c}\,t_1) \\ x_2 &= \frac{1}{\sqrt{1-\frac{v^2}{c^2}}}(c-v)t_1. \end{aligned} \tag{1.32}$$

and therefore also $x_2 = ct_2$. So all observers moving with respect to each other at constant velocity find the speed of the signal to be equal to c. In particular, there is no reference frame where the signal has zero velocity. A reference frame in which a given object has zero velocity is called the *rest frame* of the object. In electromagnetism we may therefore say that an electromagnetic wave has no rest frame.

1.3 Clock synchronization and the Lorentz transformation

The possibility of an universal speed c was an very interesting conclusion in our derivation in the previous Section. However, this conclusion was based on an, at first sight, rather innocent assumption in the derivation. The presumption was the existence of a well-defined space-time coordinate system (t, x, y, z) for a given observer in which we can assign a unique well-defined time to every spatial point. This assumption is, for instance, used when we say that the events (t, x_1, y_1, z_1) and (t, x_2, y_2, z_2) are simultaneous events that happen at the same time t at different spatial positions in our reference system. How do we know that this is true? To answer this question we must have an experimental way to verify simultaneity of distant events, otherwise the statement has no physical meaning. But what do we mean by simultaneity?
After some thought we come to the conclusion that this a matter of definition, but that some definitions are more useful than others. Let us start with a simple example. We consider an observer Ed [2] on Earth with a clock that registers local time t_{E}. This observer receives radio signals from various spacecrafts in the Solar System. A spacecraft orbiting Saturn at spatial coordinates (x_S, y_S, z_S) records a local event (for instance it took a picture of a comet hitting

[2] In general we will use different letters A,B,C,D,V,W etc. to denote different observers. However, sometimes to be more concrete and personal we will give names to these letters such as Alice, Bob, Charlotte, Dilbert, Viivi, Wagner etc.

Saturns atmosphere) and sends out a radio signal with information that is received by Ed at $t_{\rm E}$ =15:00. Five minutes later at $t_{\rm E}$ =15:05 he receives a second radio signal from another spacecraft orbiting Mars at coordinates (x_M, y_M, z_M) containing information on another local event. How should Ed give a time coordinate to these two events? He could simply decide to label the events by $(15{:}00, x_S, y_S, z_S)$ and $(15{:}05, x_M, y_M, z_M)$. In this way Ed can assign unique coordinates $(t_{\rm E}, x, y, z)$ to any event in the Solar System and in this way build a space-time coordinate system. He can then define that two events at different spatial positions are simultaneous if they have the same $t_{\rm E}$ coordinate. Although this is a well-defined prescription for building a space-time frame it is not very useful in practice. First of all, we realize that the simultaneity of two events in this way depends on the spatial location of the arbitrarily chosen origin (in our case the Earth) with respect to the events. Secondly, if the laws of physics are expressed in these space-time coordinates they attain a very complicated form. Even simple uniform motion becomes involved in these coordinates (try it yourself). We therefore do not prefer to use Ed's coordinate system. Which is then the preferred definition of simultaneity in which the laws of physics attain its simplest form?

This was discussed very clearly in Einstein's original work [1] and we will closely follow his presentation. The procedure that we will describe is commonly known as the *Einstein synchronization* of clocks. The spatial structure of our reference system can be set up without problem using standard measuring rods as described in the beginning of Section 1.1 (the underlying assumption is physical existence of Euclidean space). We restrict ourselves to reference frames in which light moves along straight lines (such reference frames are called inertial frames and will be discussed in much more detail in the next Chapter). Since the shape of the path is a purely spatial concept this requirement can be checked without the need to consider clocks. Let us therefore consider the more involved temporal structure next. We, for simplicity, consider one spatial dimension. It is clear how to measure time exactly for a local observer A (Alice) at a point $x_{\rm A}$ in the coordinate system. We can simply read off a given clock that is placed at rest at position $x_{\rm A}$. An observer B (Bob) can do the same with a local clock placed at position $x_{\rm B}$. We have therefore an "A time" and a "B time". The question is then how A and B can agree on a common definition of time. In other words, how can they properly synchronize[3] their clocks? This is solved by an operational definition which has therefore a direct experimental physical realization. Let us send a light signal from a point $x_{\rm A}$ to point $x_{\rm B}$ and let the light signal be reflected back at $x_{\rm B}$ towards $x_{\rm A}$, i.e.

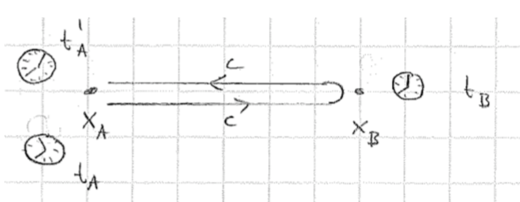

Figure 1.5: Einstein synchronization of clocks

If the signal is sent and returned in $x_{\rm A}$ at times $t_{\rm A}$ and $t'_{\rm A}$ and received in point $x_{\rm B}$ at local time $t_{\rm B}$ we *define* clocks at $x_{\rm A}$ and $x_{\rm B}$ to be synchronized when

$$t'_{\rm A} - t_{\rm B} = t_{\rm B} - t_{\rm A}$$

or equivalently when

$$t_{\rm B} = \frac{1}{2}(t_{\rm A} + t'_{\rm A}).$$

[3]Of course, "synchronous" is just a fancy way of saying "same time", synchronization is therefore just an agreement on what is meant by simultaneity.

So observer B can adjust his clock such that this relation is satisfied. More concretely, if Alice sends and receives the signal at the times 11:58 and 12:02 on her local clock and Bob receives the signal on his local clock at 12:01 then Bob concludes that his clock was running ahead by one minute and adjusts it such that the reception of the signal on his newly synchronized clock happened at 12:00 [4]. After this synchronization procedure has been carried out Alice and Bob agree by definition that an event that happened at $t_{\rm A}$ in $x_{\rm A}$ and another event that happened at $t_{\rm B}$ in $x_{\rm B}$ are to be called *simultaneous* whenever $t_{\rm A} = t_{\rm B}$.

Let us again summarize the procedure. If an observer O at the origin of the reference system places a mirror at position P at a certain distance from the origin and sends out a light signal at local time t_0 that is reflected back from the mirror and received back at local time t_1 then the local time $t_{\rm P}$ on the clock at P where the light is reflected at the mirror is defined to be

$$t_{\rm P} = \frac{1}{2}(t_0 + t_1). \tag{1.33}$$

Two events at two different spatial points P and Q in the reference frame are called simultaneous when $t_P = t_Q$. Now that we have a frame of local synchronized clocks as well as an Euclidean spatial frame we can talk about motion. In particular, if we repeat Alice and Bob's synchronization procedure for different position in our reference frame we find that $r_{\rm AB}$ is the distance between A and B we find every time that

$$c = \frac{r_{\rm AB}}{t_{\rm B} - t_{\rm A}}$$

is the same universal constant which we call light speed[5]. We can therefore, instead of using measuring sticks, also use light to define distances in our coordinate frame. If an observer O in the origin ($x = 0$) of the reference system sends out a light signal at time $t_{\rm O}$ which is received by a local observer P at time t_P then the distance of P from the origin is then given by

$$x_{\rm P} = c\,(t_{\rm P} - t_{\rm O}). \tag{1.34}$$

(more precisely we can define a coordinate $x_{\rm P} = \pm c\,(t_{\rm P} - t_{\rm O})$ at a distance $|x_{\rm P}|$ removed from the origin dependent on whether P lies on the positive or negative x-axis). In this way observer O can construct experimentally the whole space-time coordinate system $(t_{\rm P}, x_{\rm P})$ for all points P only using clocks and light sources. An object is then defined to move at uniform speed v with respect to this reference system whenever $v = (x_{\rm P} - x_{\rm Q})/(t_{\rm P} - t_{\rm Q})$ is the same constant for all positions $x_{\rm P}$ and $x_{\rm Q}$ that the object passes. We find for our Einstein-synchronized reference frame that isolated material bodies that do not interact with other bodies move according to the simple law

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

where t is the time displayed by a local Einstein-synchronized clock at position $\mathbf{x}$ for a constant velocity vector $\mathbf{v}$. Any other synchronization scheme (for an extensive discussion see [3]) will lead to less simple laws. As a final remark we note that our synchronization applies also to the Newtonian limit by taking the limit $c \to \infty$ in which case distant clocks are trivially synchronized by infinite speed signals.

With these premises we can now derive the Lorentz transformation the standard way [2] in which the basic postulate is the universality of the speed of light independent of the state of motion of the light source. More specifically the assumptions are

[4] Bob is two light minutes away from Alice. If Alice were on Earth then Bob could roughly be at the orbit of planet Venus around the Sun.

[5] This property was tested in the famous Michelson and Morley experiment which aimed to find out if the motion of the Earth has any influence on the light speed.

1. Once Einstein synchronization has been performed in any inertial frame, the linear space-time coordinate transformations between any two frames have to be symmetric and dependent on the relative velocity of the two frames alone.

2. In any frame, once Einstein synchronization has been performed, the velocity of light is equal to c along any path, independent of the state of motion of the emitting body.

The first requirement is equivalent to condition (1.8) of the previous Section. The second requirement is an experimentally well verified fact. We consider again two systems O_1 and O_2 that use coordinates (t, x) and (t', x') constructed using Einstein synchronization as described above. Let O_2 move at uniform relative speed v with respect to O_1. We further assume that the origins of both coordinate systems coincide at $x = t = 0$ and $x' = t' = 0$. Then, according to assumption 2 in these coordinate systems the identities

$$x - c\,t = 0 \tag{1.35}$$
$$x' - c\,t' = 0 \tag{1.36}$$

are valid. The space-time points that satisfy (1.35) must at the same time satisfy (1.36). Since we look for a linear relation between the coordinate systems this is generally true whenever

$$x' - c\,t' = \lambda\,(x - c\,t) \tag{1.37}$$

where λ is a constant, since the vanishing of either side of the equation implies the vanishing of the other side. A completely analogous reasoning applies to light signals that move in the negative direction along the x-axis. We therefore have also

$$x' + c\,t' = \mu\,(x + c\,t). \tag{1.38}$$

If we add and subtract Eqs.(1.37) and (1.38) and define

$$a = \frac{1}{2}(\lambda + \mu) \qquad b = \frac{1}{2}(\lambda - \mu)$$

we find

$$x' = ax - bct \tag{1.39}$$
$$ct' = act - bx. \tag{1.40}$$

It remains to find the constants a and b. For the origin of O_2 we have $x' = 0$ and therefore from Eq.(1.39) it follows that

$$x = \frac{bct}{a}.$$

Since system O_2 moves with velocity v with respect to O_1 the position of the origin of O_2 is given by $x = vt$ and therefore

$$v = \frac{bc}{a}.$$

Then $b = av/c$ and Eqs.(1.39) and (1.40) become

$$x' = a(x - vt) \tag{1.41}$$
$$t' = a(t - \frac{vx}{c^2}). \tag{1.42}$$

Let us now consider all events that happen at time $t = 0$ in system O_1. Then those events happen at position

$$x' = ax$$

in system O_2. In particular the points $x'=1$ and $x'=0$ are separated by a distance

$$\Delta_1 = \frac{1}{a}$$

as viewed by O_1. Let us now make the similar snapshot but now viewed from system O_2. We take all events that happen at time $t'=0$ in system O_2. Then Eq.(1.42) tells us that $t = vx/c^2$ which inserted into Eq.(1.41) yields

$$x' = ax(1-\frac{v^2}{c^2})$$

In particular the points $x=1$ and $x=0$ are separated by a distance

$$\Delta_2 = a(1-\frac{v^2}{c^2})$$

as viewed by O_2. Now we invoke assumption 1 that coordinate frames that move uniformly with respect to each other are completely equivalent for the description of physical phenomena. We must therefore have that $\Delta_1 = \Delta_2$ and hence

$$a^2 = \frac{1}{1-\frac{v^2}{c^2}} \quad \Rightarrow \quad a = \frac{\pm 1}{\sqrt{1-\frac{v^2}{c^2}}}$$

We see, however, from Eqs.(1.41) and (1.42) that for $v=0$ we must have $a=1$ and therefore the negative solution is excluded. We then obtain

$$\begin{aligned} x' &= \frac{x-vt}{\sqrt{1-\frac{v^2}{c^2}}} \\ t' &= \frac{t-\frac{vx}{c^2}}{\sqrt{1-\frac{v^2}{c^2}}} \end{aligned}$$

which is exactly the Lorentz transformation that we derived before. The theory that Einstein developed on the basis of these coordinate transformations has become famous as the Special theory of Relativity. In the following Sections we explore some of its consequences.

1.4 Lorentz symmetry and its consequences

So far we considered only motion in one spatial dimension. However, nothing essential changes by going to three spatial dimensions.

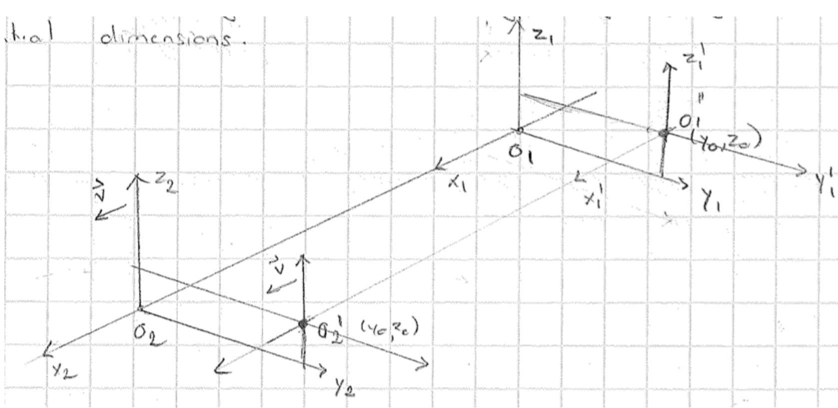

Figure 1.6: Relative motion of coordinate frames

It is clear from the symmetry of the motion that we considered that an event that happens at the x_1-axis in system O_1 happens on the x_2-axis in system O_2. It can not be mapped outside the x_2-axis due to the rotational symmetry around the common x-axes, as this would break the symmetry, i.e.

$$(t_1, x_1, 0, 0) \to (t_2, x_2, 0, 0) \tag{1.43}$$

However, the choice of the origin in the planes $x_1 = 0$ and $x_2 = 0$ is completely arbitrary as there is no preferred origin. We might have chosen an alternative coordinate system for O_1 where

$$\begin{aligned} x_1' &= x_1 \\ y_1' &= y_1 + y_0 \\ z_1' &= z_1 + z_0 \end{aligned}$$

and similarly for system O_2 and conclude Eq. (1.43) as well in our primed coordinate systems. However, in our original coordinate system this amounts to

$$(t_1, x_1, y_0, z_0) \to (t_2, x_2, y_0, z_0) \tag{1.44}$$

i.e. $y_1 = y_2$ and $z_1 = z_2$. The generalization of the Lorentz transformation to three spatial dimensions is therefore given by

$$\begin{pmatrix} t_2 \\ x_2 \\ y_2 \\ z_2 \end{pmatrix} = A(v) \begin{pmatrix} t_1 \\ x_1 \\ y_1 \\ z_1 \end{pmatrix}$$

where

$$A(v) = \begin{pmatrix} \gamma & -\gamma \frac{v}{c^2} & 0 & 0 \\ -\gamma v & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{1.45}$$

and we defined

$$\gamma = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} \tag{1.46}$$

From Eq.(1.45) we can deduce the transformation law for velocities. Let a particle trajectory in systems O_1 and O_2 by given by

$$\begin{aligned} \mathbf{x}_1(t_1) &= (x_1(t_1), y_1(t_1), z_1(t_1)) \\ \mathbf{x}_2(t_1) &= (x_2(t_2), y_2(t_2), z_2(t_2)) \end{aligned}$$

The velocities of the particles is then given by

$$\mathbf{u}_1(t_1) = \frac{d\mathbf{x}_1}{dt_1}(t_1) \quad \text{and} \quad \mathbf{u}_2(t_2) = \frac{d\mathbf{x}_2}{dt_2}(t_2)$$

Since

$$\begin{aligned} x_2 &= \gamma(t_1 - \frac{v}{c^2} x_1) \\ y_2 &= y_1 \\ z_2 &= z_1 \end{aligned}$$

it follows that

$$\begin{aligned} u_{2,\mathrm{x}} &= \frac{dx_2}{dt_2} = \gamma\left(\frac{dt_1}{dt_2} - \frac{v}{c^2}\frac{dx_1}{dt_2}\right) = \gamma\left(1 - \frac{v}{c^2}u_{1,\mathrm{x}}\right)\frac{dt_1}{dt_2} \\ u_{2,\mathrm{y}} &= \frac{dy_2}{dt_2} = \frac{dy_1}{dt_1}\frac{dt_1}{dt_2} = u_{1,\mathrm{y}}\frac{dt_1}{dt_2} \\ u_{2,\mathrm{z}} &= \frac{dz_2}{dt_2} = \frac{dz_1}{dt_1}\frac{dt_1}{dt_2} = u_{1,\mathrm{z}}\frac{dt_1}{dt_2} \end{aligned}$$

Since, by the inverse Lorentz transformation

$$t_1 = \gamma(t_2 + \frac{v}{c^2}x_2)$$

it follows that

$$\frac{dt_1}{dt_2} = \gamma(1 + \frac{v}{c^2}u_{2,\mathrm{x}})$$

Collecting our results we therefore obtain after some rewriting

$$u_{1,\mathrm{x}} = \frac{u_{2,\mathrm{x}} + v}{1 + \frac{v}{c^2}u_{2,\mathrm{x}}} \tag{1.47}$$

$$u_{1,\mathrm{y}} = u_{2,\mathrm{y}}\frac{(1 - \frac{v^2}{c^2})^{\frac{1}{2}}}{1 + \frac{v}{c^2}u_{2,\mathrm{x}}} \tag{1.48}$$

$$u_{1,\mathrm{z}} = u_{2,\mathrm{z}}\frac{(1 - \frac{v^2}{c^2})^{\frac{1}{2}}}{1 + \frac{v}{c^2}u_{2,\mathrm{x}}} \tag{1.49}$$

For the special case that $u_{2,\mathrm{y}} = u_{2,\mathrm{z}} = 0$ we recover Eq.(1.31) that we derived before. Taking the limit $c \to \infty$ gives the familiar Newtonian limit. Let us now deduce a number of well-known consequences of the Lorentz transformation. For simplicity we again consider motion in one spatial dimension. From Eqs.(1.27) and (1.28) we see that

$$\begin{aligned} -c^2t_2^2 + x_2^2 &= -\gamma^2(ct_1 - \frac{v}{c}x_1)^2 + \gamma^2(x_1 - vt_1)^2 \\ &= -\gamma^2[(c^2 - v^2)t_1^2 - (1 - \frac{v^2}{c^2})x_1^2] = -c^2t_1^2 + x_1^2 \end{aligned} \tag{1.50}$$

The equation

$$-c^2t^2 + x^2 = \mathrm{constant}$$

represents a hyperbola in the (ct, x)-plane.

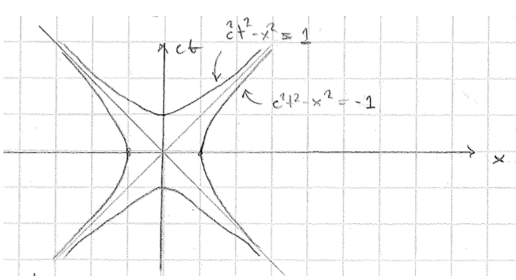

Figure 1.7: Hyperbola in the (ct, x)-plane. Note that it is convention to take the time axis to be the vertical axis.

We therefore see that a Lorentz transformation maps a point on a hyperbola to a point on the same hyperbola. In the Fig.1.7 we have drawn the hyperbola for $c^2t^2 - x^2 = \pm 1$. It is convenient to denote $\tau = ct$ such that the Lorentz transformation attains the form

$$\tau_2 = \gamma(\tau_1 - \frac{v}{c}x_1) \tag{1.51}$$

$$x_2 = \gamma(x_1 - \frac{v}{c}\tau_1) \tag{1.52}$$

The set of points that occur at equal time $\tau_2 = K =$ constant in system O_2 correspond according to (1.51) to the set of lines

$$\tau_1 = \frac{v}{c}x_1 + \frac{K}{\gamma} \tag{1.53}$$

in the (τ_1, x_1)-plane. Similarly the set of space-time points at equal position $x_2 = K =$ constant in system O_2 is according to (1.52) given by the lines

$$\tau_1 = \frac{c}{v}x_1 - \frac{c}{v}\frac{K}{\gamma} \tag{1.54}$$

i.e.

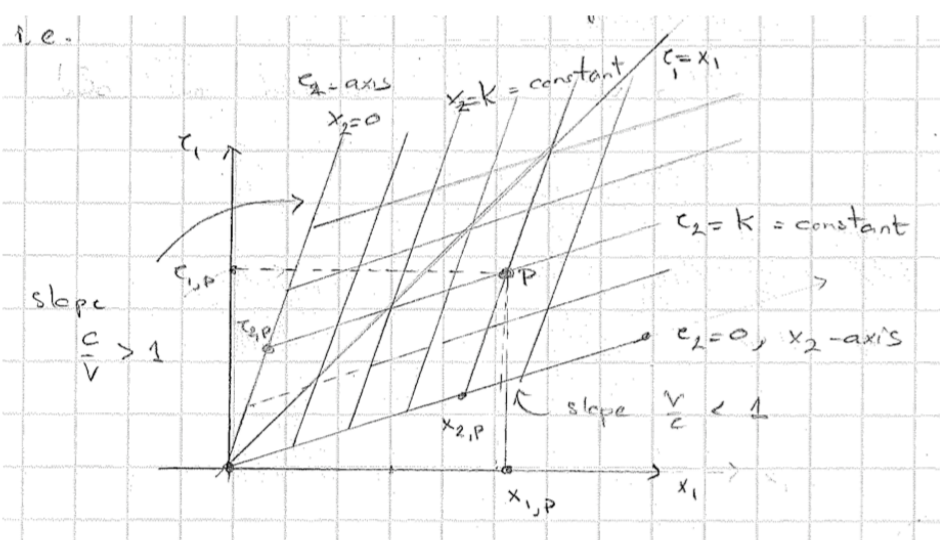

Figure 1.8: Relation between coordinates

The points $(\tau_2 = 0, x_2)$ represent the x_2-axis of system O_2 and the points $(\tau_2, x_2 = 0)$ represent the τ_2 axis of this system. The coordinates of a space-time point $(\tau_{1,\mathrm{P}}, x_{1,\mathrm{P}})$ in the (τ_1, x_1)-plane can thus be transformed to coordinates $(\tau_{2,\mathrm{P}}, x_{2,\mathrm{P}})$ in system O_2 by parallel projection along the lines $\tau_2 =$ constant and $x_2 =$ constant. However, the length scale on the O_2-axes is not the same as that of the O_1-axes. The unit point $(\tau_2, x_2) = (0, 1)$ lies on the hyperbola $\tau_1^2 - x_1^2 = -1$ (see Eq.(1.50)) whereas the unit point $(\tau_2, x_2) = (1, 0)$ lies on the hyperbola $\tau_1^2 - x_1^2 = 1$. We thus have

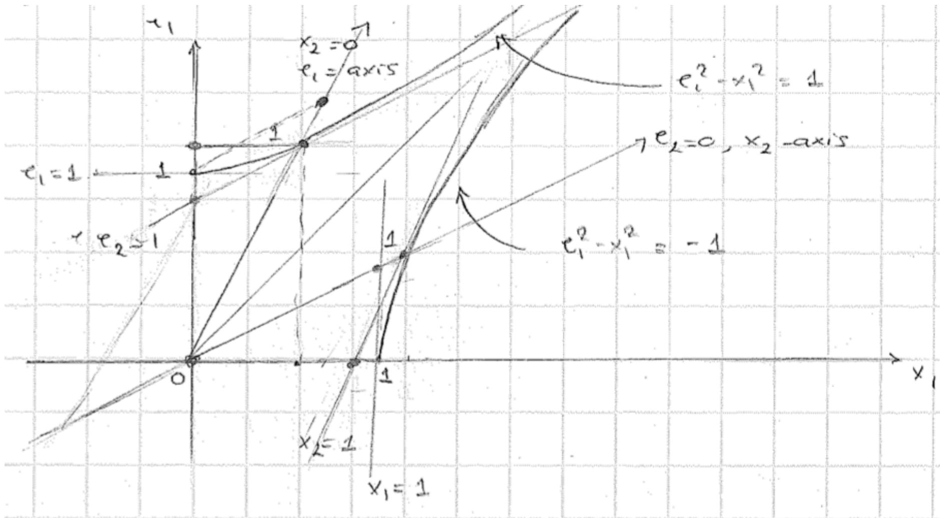

Figure 1.9: Intersecting hyperbolas

where the intersection of the hyperbola $\tau_1^2 - x_1^2 = -1$ with the x_2-axis $\tau_1 = vx_1/c$ corresponds to the unit on the x_2-axis. One can check that the tangent to the hyperbola in this point has slope c/v and is parallel to the τ_2-axis. Similarly the intersection of hyperbola $\tau_1^2 - x_1^2 = 1$ with the τ_2-axis $\tau_1 = cx_1/v$ yields the unit point on the τ_2-axis. The tangent to the hyperbola in this point has slope v/c and is parallel to the x_2-axis.

A number of interesting facts can be read off this diagram:

1. The axes $x_2 = 1$ and $x_2 = 0$ intersect the x_1-axis at equal time $\tau_1 = 0$ in system O_1 at a distance less than one, i.e. the length of a measuring rod at rest in O_2 seems smaller in system O_1. Let us check this explicitly. If the end points of a rod at rest in O_2 are $x_{2,A}$ and $x_{2,B}$ then at equal time $t_1 = t$ in O_1

$$\begin{aligned} x_{2,A} &= \gamma(x_{1,A} - vt) \\ x_{2,B} &= \gamma(x_{1,B} - vt) \end{aligned}$$

Subtraction of these two equations then gives

$$x_{2,A} - x_{2,B} = \gamma(x_{1,A} - x_{1,B}) \Rightarrow |x_{1,A} - x_{1,B}| = \sqrt{1 - \frac{v^2}{c^2}}|x_{2,A} - x_{2,B}| < |x_{2,A} - x_{2,B}|$$

This phenomenon is known as *Lorentz contraction*.

2. We see from figure 1.9 that a clock at rest in system O_2 which records a value one, records a value larger than one as seen from system O_1, i.e. with respect to O_1 the moving clock in system O_2 runs slower. Let us again check this explicitly. The points $(\tau_2, x_2) = (0,0)$ and $(\tau_2, x_2) = (1,0)$ correspond to $(\tau_1, x_1) = (0,0)$ and

$$\begin{pmatrix} \tau_1 \\ x_1 \end{pmatrix} = \gamma \begin{pmatrix} 1 & \frac{v}{c} \\ \frac{v}{c} & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} \gamma \\ \frac{v}{c}\gamma \end{pmatrix}$$

Clearly the time elapse between the two space-time points in O_1 is γ whereas it is 1 in O_2 and $\gamma > 1$. From the figure we see that reciprocally a similar time dilatation is observed from system O_2 that sees the clocks in system O_1 run slower.

A different kind of dilatation is observed when system O_1 receives a periodic signal of light flashes emitted from system O_2. In terms of a space-time diagram the situation is as follows.

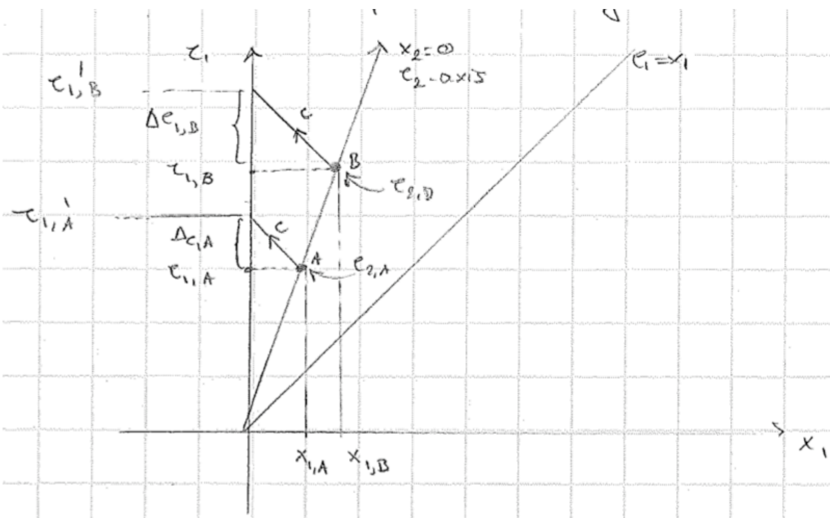

Figure 1.10: Description of the Doppler-effect

System O_2 emits a light signal in $x_2 = 0$ at time $\tau_{2,\mathrm{A}}$. The space-time coordinates of this point in system O_1 are

$$\begin{pmatrix} \tau_{1,\mathrm{A}} \\ x_{1,\mathrm{A}} \end{pmatrix} = \gamma \begin{pmatrix} 1 & \frac{v}{c} \\ \frac{v}{c} & 1 \end{pmatrix} \begin{pmatrix} \tau_{2,\mathrm{A}} \\ 0 \end{pmatrix} = \begin{pmatrix} \gamma\,\tau_{2,\mathrm{A}} \\ \frac{v}{c}\gamma\,\tau_{2,\mathrm{A}} \end{pmatrix}$$

The light signal as observed in O_1 takes time $\Delta t_{1,\mathrm{A}} = x_{1,\mathrm{A}}/c$ to travel from the point $x_{1,\mathrm{A}}$ to the origin of system O_1, i.e.

$$\Delta\tau_{1,\mathrm{A}} = c\,\Delta t_{1,\mathrm{A}} = x_{1,\mathrm{A}} = \frac{v}{c}\gamma\,\tau_{2,\mathrm{A}}.$$

The light signal will therefore be received on the τ_1-axis at time point

$$\tau_{1,\mathrm{A}'} = \tau_{1,\mathrm{A}} + \Delta\tau_{1,\mathrm{A}} = \gamma\,\tau_{2,\mathrm{A}} + \gamma\,\frac{v}{c}\,\tau_{2,\mathrm{A}} = \gamma(1+\frac{v}{c})\tau_{2,\mathrm{A}} = \sqrt{\frac{1+\frac{v}{c}}{1-\frac{v}{c}}}\,\tau_{1,\mathrm{A}}$$

Subsequently a light signal is emitted a time $\tau_{2,\mathrm{B}}$ in system O_2. Exactly the same calculation as before applies to point B in the space-time diagram. So we have

$$\tau_{1,\mathrm{B}'} - \tau_{1,\mathrm{A}'} = \sqrt{\frac{1+\frac{v}{c}}{1-\frac{v}{c}}}\,(\tau_{2,\mathrm{B}} - \tau_{2,\mathrm{A}})$$

or

$$t_{1,\mathrm{B}'} - t_{1,\mathrm{A}'} = \sqrt{\frac{1+\frac{v}{c}}{1-\frac{v}{c}}}\,(t_{2,\mathrm{B}} - t_{2,\mathrm{A}}) \tag{1.55}$$

So, depending on whether v is positive or negative, the time difference between the arrival of the signals is longer or shorter in O_1 than the time-difference between the emissions in O_2. The frequency of arrival of the signals is then $\nu_1 = 1/(t_{1,\mathrm{B}'} - t_{1,\mathrm{A}'})$ whereas the frequency of emission is $\nu_2 = 1/(t_{2,\mathrm{B}} - t_{2,\mathrm{A}})$ which according to (1.55) yields

$$\nu_1 = \sqrt{\frac{1-\frac{v}{c}}{1+\frac{v}{c}}}\,\nu_2 \tag{1.56}$$

This is the formula for the Doppler-shift. For $v > 0$ we have $\nu_1 < \nu_2$ and system O_1 observes a redshift whereas for $v < 0$ we have $\nu_1 > \nu_2$ and we observe a blueshift.
We finally discuss the issue of clock synchronization and simultaneity using the following example

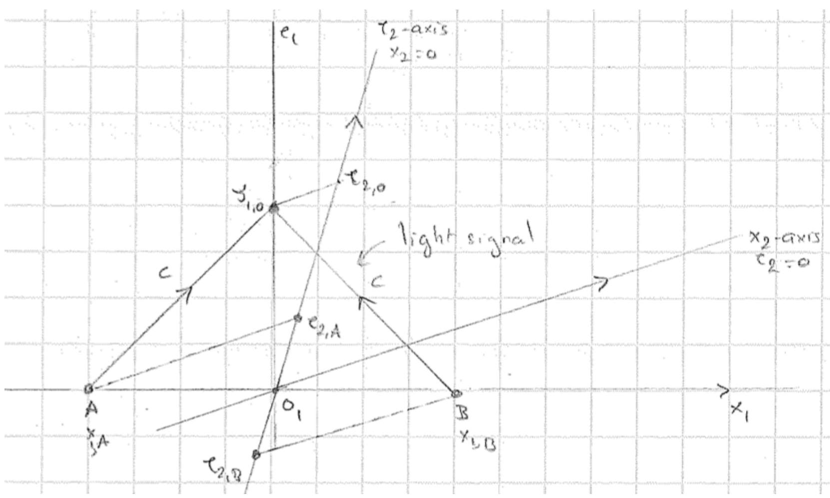

Figure 1.11: Relativity of simultaneity

Two clocks A and B are positioned at points $x_{1,\mathrm{A}} = -x_{1,\mathrm{B}}$ (see figure). At time $\tau_1 = 0$ both clocks simultaneously (in system O_1) emit a light signal towards the origin of system O_1 (containing for instance a photographic image of the face of the clock). Both signals are received in the origin at time $\tau_{1,\mathrm{O}}$ when observer O_1 looks at the images of both clocks and is happy to see that they record both the same time and are properly synchronized. Now from the viewpoint of O_2 the light signals arrived at the same time $\tau_{2,\mathrm{O}}$ in the origin of O_1 but they were, however, not emitted at the same time. The signal from clock B left at a time $\tau_{2,\mathrm{B}}$ before the time $\tau_{2,\mathrm{A}}$ at which the signal from clock A left. For O_2 the light emissions from clocks A and B were not simultaneous. Since the readings of both clocks by O_1 was the same, observer O_2 therefore concludes that clocks A and B were not properly synchronized.

1.5 World lines and proper time

In the previous Section we of drew the trajectories of light signals in a space-time diagram. In general we can draw the trajectory of a particle that moves with a non-uniform velocity. This is called a *world line*.

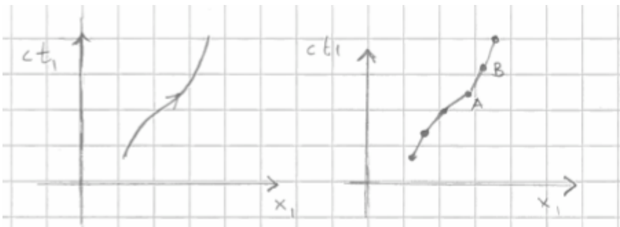

Figure 1.12: World line and its approximation with linear segments.

Since the particle can never move faster than c there is a limit on the slope of the tangent to the world line. We must have that

$$\frac{d\tau_1}{dx_1} = \frac{c}{v(t_1)} > 1$$

So the slope of the tangent to a world line must always be larger than 1. The next issue we want to address is what the reading of the local time is on a clock traveling with the particle. Since the motion is not linear we will first approximate the curve with linear segments and take the limit of infinitesimal segments at the end (see Fig. 1.12). In fact, the segmentation is not just a mathematical procedure but rather close to the true physical situation since it is experimentally very hard to measure a continuous curve. In practice the observer measures the position of the particle at a few points and registers the local time at these positions, such that we have a table of points (t_j, x_j). A linear interpolation between these points then produces the segmented world line of Fig. 1.12.

We saw in Eq.(1.50) that a Lorentz transformation maps points on a hyperbola to the same hyperbola. A similar relation is true for the difference of the coordinates of two space-time events A and B:

$$-c^2(t_{1,\mathrm{A}} - t_{1,\mathrm{B}})^2 + (x_{1,\mathrm{A}} - x_{1,\mathrm{B}})^2 = -c^2(t_{2,\mathrm{A}} - t_{2,\mathrm{B}})^2 + (x_{2,\mathrm{A}} - x_{2,\mathrm{B}})^2 \tag{1.57}$$

where $(t_{1,\mathrm{A}}, x_{1,\mathrm{A}})$ is the space-time event A in system O_1 and $(t_{2,\mathrm{A}}, x_{2,\mathrm{A}})$ is the same space-time event in system O_2, and similarly for event B. This equation follows immediately from the linearity of the Lorentz transformation, i.e. if for $\mathbf{x} = (\tau, x)$

$$\begin{aligned}\mathbf{x}_{2,\mathrm{A}} &= A(v)\mathbf{x}_{1,\mathrm{A}} \\ \mathbf{x}_{2,\mathrm{B}} &= A(v)\mathbf{x}_{1,\mathrm{B}}\end{aligned}$$

then also $\mathbf{x}_{2,\mathrm{A}} - \mathbf{x}_{2,\mathrm{B}} = A(v)(\mathbf{x}_{1,\mathrm{A}} - \mathbf{x}_{1,\mathrm{B}})$ and Eq. (1.57) can be derived exactly as in Eq.(1.50). If we denote

$$\begin{aligned}\Delta t_1 &= t_{1,\mathrm{A}} - t_{1,\mathrm{B}} \\ \Delta x_1 &= x_{1,\mathrm{A}} - x_{1,\mathrm{B}}\end{aligned}$$

and similarly in system O_2 we can write Eq.(1.57) as

$$-c^2\,\Delta t_1^2 + \Delta x_1^2 = -c^2\,\Delta t_2^2 + \Delta x_2^2$$

Let these differences describe events on the segmented world line of the particle. We take the case that system O_2 travels with the particle such that $\Delta x_2 = 0$ in this reference frame. We can then write

$$-c^2\,\Delta t_1^2 + \Delta x_1^2 = -c^2\,\Delta t_2^2$$

or equivalently

$$\Delta t_2 = \sqrt{1 - \frac{1}{c^2}\left(\frac{\Delta x_1}{\Delta t_1}\right)^2}\,\Delta t_1.$$

If we make the space-time differences very small then

$$\frac{\Delta x_1}{\Delta t_1} \approx \frac{dx_1}{dt_1}(t_1) = v(t_1)$$

is approximately the instantaneous velocity of the particle at time t_1 as observed in system O_1. In system O_2, in which the particle is approximately at rest, a time difference

$$\Delta t_2 \approx \sqrt{1 - \frac{v(t_1)^2}{c^2}}\,\Delta t_1.$$

has elapsed between the space-time events. This equation becomes exact when we write

$$\frac{dt_2}{dt_1} = \sqrt{1 - \frac{v(t_1)^2}{c^2}} \tag{1.58}$$

Integrating this equation then gives

$$t_2(t) - t_2(t_0) = \int_{t_0}^{t} dt_1 \sqrt{1 - \frac{v(t_1)^2}{c^2}}$$

The left hand side of this equation describes the time elapsed on a clock that travels with the particle between space-time points $(t, x(t))$ and (t_0, x_0) as observed by O_1. This time is usually referred to as the *proper time* of the particle and is usually denoted by τ (not to be confused with $\tau = ct$ used earlier). So we have

$$\tau(t) = \tau(t_0) + \int_{t_0}^{t} dt_1 \sqrt{1 - \frac{v(t_1)^2}{c^2}} \tag{1.59}$$

As is physically clear the proper time is a relativistic invariant independent of the frame used to evaluate Eq.(1.59), i.e.

$$\begin{aligned} \tau(t) - \tau(t_0) &= \int_{t_0}^{t} dt_1 \sqrt{1 - \frac{1}{c^2}\left(\frac{dx_1}{dt_1}\right)^2} \\ &= \int_{t_0'}^{t'} dt_2 \sqrt{1 - \frac{1}{c^2}\left(\frac{dx_2}{dt_2}\right)^2} = \tau(t') - \tau(t_0') \end{aligned} \tag{1.60}$$

as explained in the diagram

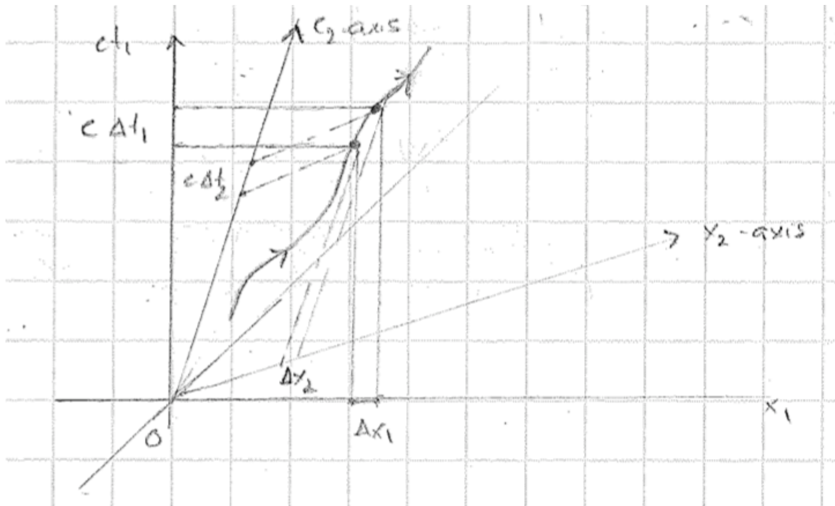

Figure 1.13: Proper time on a world line

where we drew the world line $(ct_1(\tau), x_1(\tau))$ parametrized by the proper time as viewed in two different systems O_1 and O_2.
It is clear that the analysis can be directly generalized to the case of three spatial dimensions. In that case the Lorentz invariant distance between space-time points is given by

$$-c^2(t_{1,\mathrm{A}} - t_{1,\mathrm{B}})^2 + (x_{1,\mathrm{A}} - x_{1,\mathrm{B}})^2 + (y_{1,\mathrm{A}} - y_{1,\mathrm{B}})^2 + (z_{1,\mathrm{A}} - z_{1,\mathrm{B}})^2 = \mathrm{constant} \quad (1.61)$$

This constant is usually denoted by $(\Delta s)^2$ in analogy with the usual Euclidean squared distance, despite the fact that it can be negative[6]. We therefore can write

$$(\Delta s)^2 = -c^2(\Delta t)^2 + (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2 \quad (1.62)$$

Very commonly an infinitesimal version of this expression is used, i.e.

$$ds^2 = -c^2 dt^2 + dx^2 + dy^2 + dz^2 \quad (1.63)$$

which is especially useful for transformations to general coordinates, such as a spherical coordinate system. For light signals we always have that $ds^2 = 0$ while motions of material points are characterized by $ds^2 < 0$. We can now repeat the derivation of the one-dimensional example. If $(t(\tau), x(\tau), y(\tau), z(\tau))$ is the world line of a particle as observed in an inertial frame and if $d\tau$ is the increase in proper time between the space-time points (t, x, y, z) and $(t + dt, x + dx, y + dy, z + dz)$ then we have

$$-c^2 d\tau^2 = -c^2 dt^2 + dx^2 + dy^2 + dz^2 \quad (1.64)$$

and hence

$$\begin{aligned}\tau(t) - \tau(t_0) &= \int_{t_0}^{t} dt_1 \sqrt{1 - \frac{1}{c^2}\left[\left(\frac{dx}{dt_1}\right)^2 + \left(\frac{dy}{dt_1}\right)^2 + \left(\frac{dz}{dt_1}\right)^2\right]} \\ &= \int_{t_0}^{t} dt_1 \sqrt{1 - \frac{v(t_1)^2}{c^2}} \quad (1.65)\end{aligned}$$

[6]In that case Δs would be imaginary. However, in practice we only need the square of the element so we never need to deal with imaginary numbers.

Since the integrand in this equation is always smaller or equal to one we have for any space-time path that

$$\tau(t) - \tau(t_0) \leq t - t_0 \tag{1.66}$$

Therefore the time passed on the local clock traveling with the particle is always smaller than the clock in the inertial frame. Therefore, when the traveling observer returns to the origin of the inertial frame and compares clocks he will find that his clock has recorded a smaller time interval. The discrepancy is real and has been verified experimentally. The reason for the difference is that the moving observer has not been moving at uniform speed. After all he had to slow down and return. The accelerations that were caused by this change in speed have caused the slow down of the moving clock as compared the clock in the inertial frame. More discussions on clock rates in non-inertial frames can be found in the next Chapter.

Chapter 2

Reference frames and coordinate invariance

Inertial frames are defined as frames in which the physical laws are invariant under rigid space and time translations as well as the orientation of the coordinate axes. We discuss Newton's law of gravitation as an example of a physical law and show how it transforms to different reference frames. We discuss the practical and conceptual difficulties with the special status of inertial frames. As an example we consider a rotating system described both using Newtonian mechanics and special relativity. Finally, we discuss the arbitrariness in defining space-time frames and the physical meaning of particle trajectories.

2.1 Inertial frames

In the previous Chapter we discussed the possible linear coordinate transformations compatible with a certain group structure and used them to describe the motion of objects without consideration of the forces that cause this motion. What is left, of course, is a description of how forces change the motion of bodies. This obviously involves a description of the reason for the change of motion as formulated in a physical law. Newton discovered the basic laws of classical mechanics as well as the law that describes the motion of bodies under the influence of their gravitational attraction. He also realized that these mechanical laws attain a simple form but only in certain privileged reference frames. These reference frames are called *inertial frames* and they are characterized by the property that they treat space and time homogeneously as well as space isotropically. This means, more specifically, that in such frames the mechanical laws (or more generally all physical laws) are invariant under translations in space and time as well as under rotation of the coordinate axes. Let us discuss this in more detail with an example.
We go back to Newtonian mechanics and consider in a given coordinate system O a number of N masses $m_i, i = 1 \dots N$ at positions given by the vectors $\mathbf{x}_i(t)$. Then Newton states that there exists a class of reference frames such that in these frames motion of the masses m_i under the influence of each others gravitational forces is given by the differential equation

$$\mathbf{F}_i(t) = m_i \frac{d^2\mathbf{x}_i}{dt^2}(t) = -G \sum_{j \neq i}^{N} m_i m_j \frac{\mathbf{x}_i(t) - \mathbf{x}_j(t)}{|\mathbf{x}_i(t) - \mathbf{x}_j(t)|^3} \tag{2.1}$$

where G is the gravitational constant and $\mathbf{F}_i$ is the force on mass m_i. Eq.(2.1) represents a set of second order differential equations that needs to be solved with the initial conditions $\mathbf{x}_i(t_0) = \mathbf{x}_{i,0}$ and $d\mathbf{x}_i(t_0)/dt = \mathbf{v}_{i,0}$. For three particles we, for instance, have

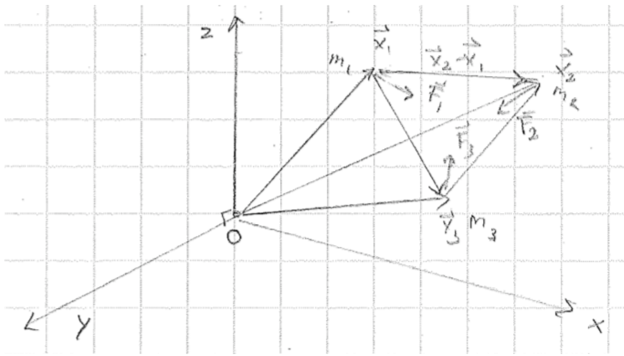

Figure 2.1: Three interacting masses

Consider now the following three operations:

$$\mathbf{x}'_i(t) = \mathbf{x}_i(t+t_0) \tag{2.2}$$
$$\mathbf{x}'_i(t) = \mathbf{x}_i(t) + \mathbf{x}_0 \tag{2.3}$$
$$\mathbf{x}'_i(t) = R\,\mathbf{x}_i(t) \tag{2.4}$$

where R in the last equation is a time-independent rotation matrix and t_0 and $\mathbf{x}_0$ are an arbitrary time and position. It is an easy exercise to check that Eq.(2.1) is invariant under these three transformations such that after each of these transformations it attains the form

$$m_i \frac{d^2\mathbf{x}'_i}{dt^2}(t) = -G \sum_{j \neq i}^{N} m_i m_j \frac{\mathbf{x}'_i(t) - \mathbf{x}'_j(t)}{|\mathbf{x}'_i(t) - \mathbf{x}'_j(t)|^3}. \tag{2.5}$$

We therefore see that the class of reference frames for which Eq.(2.1) is valid treats this physical law in a way that is homogeneous in space and time, as well as isotropic in space. As mentioned above, such reference frames are called inertial frames. The next question is then how to identify an inertial frame in practice. Let us first start by showing that a reference frame moving at uniform velocity with respect to a given inertial frame is another inertial frame.
Consider a second observer O' moving with respect to O at constant velocity $\mathbf{v}$. If the coordinate axes of O and O' are parallel and coincide at $t = 0$ then the coordinates in O and O' are related by the Galilean transformation

$$\begin{aligned} \mathbf{x}' &= \mathbf{x} - \mathbf{v}\,t \\ t' &= t \end{aligned}$$

It is immediately clear from this transformation that difference vectors are invariant under this transformation, i.e.

$$\mathbf{x}'_1 - \mathbf{x}'_2 = (\mathbf{x}_1 - \mathbf{v}\,t) - (\mathbf{x}_1 - \mathbf{v}\,t) = \mathbf{x}_1 - \mathbf{x}_2$$

and furthermore

$$\frac{d^2\mathbf{x}'}{dt^2} = \frac{d^2}{dt^2}(\mathbf{x} - \mathbf{v}\,t) = \frac{d^2\mathbf{x}}{dt^2}.$$

These equations imply that with respect to system O' Newton's equations (2.1) again attain the form of Eq.(2.5). It is therefore clear that the new frame is again an inertial frame. Since Newton's equations have, after the Galilean transformation, exactly the same form as Eq.(2.1) we say say that the equations are Galilean invariant. This means physically that if observer O' puts the masses m_i at the same initial positions and gives them the same initial velocities as observer O, i.e. $\mathbf{x}'_i(t_0) = \mathbf{x}_{i,0}$ and $d\mathbf{x}'_i(t_0)/dt = \mathbf{v}_{i,0}$ the solutions will clearly be identical.

In other words if O and O' perform the same experiment the outcomes will be the same and therefore observers O and O' are physically equivalent. In the derivation of this result the form of the interactions on the right hand side of Eq.(2.1) does not play any role. What is important, however, is that the interactions only depends on the difference vectors $\mathbf{x}_i - \mathbf{x}_j$ and respect isotropy. More generally we can instead of Eq.(2.1) write

$$m_i \frac{d^2\mathbf{x}_i}{dt^2}(t) = -\frac{\partial V}{\partial \mathbf{x}_i}(\mathbf{x}_1, \ldots, \mathbf{x}_N) \tag{2.6}$$

where

$$V(\mathbf{x}_1, \ldots, \mathbf{x}_N) = \sum_{i>j}^{N} v_{ij}(\mathbf{x}_i - \mathbf{x}_j)$$

is a potential only dependent on the difference vectors $\mathbf{x}_i - \mathbf{x}_j$ (it also does not need to be a sum of two-body potentials v_{ij} as above, but this is the most common physical situation). The choice

$$v_{ij}(\mathbf{r}) = -\frac{G m_i m_j}{|\mathbf{r}|}$$

then gives back Eq.(2.1). We can say that Eq.(2.6) is Galilean invariant. There is, however, no invariance of Eq.(2.1) or (2.6) under more general transformations. Suppose now that system O' differs from O by a combination of a time-dependent translation and rotation, i.e. instead of the Galilean transformation we have

$$\mathbf{x}(t) = R(t)\mathbf{x}'(t) + \mathbf{r}(t)$$

where $R(t)$ is a time-dependent rotation matrix and $\mathbf{r}(t)$ a time-dependent translation vector. Then

$$\begin{aligned} \frac{d\mathbf{x}}{dt} &= \frac{dR(t)}{dt}\mathbf{x}'(t) + R(t)\frac{d\mathbf{x}'}{dt} + \frac{d\mathbf{r}}{dt} \\ \frac{d^2\mathbf{x}}{dt^2} &= R(t)\frac{d^2\mathbf{x}'}{dt^2} + 2\frac{dR(t)}{dt}\frac{d\mathbf{x}'}{dt} + \frac{d^2R(t)}{dt^2}\mathbf{x}'(t) + \frac{d^2\mathbf{r}}{dt^2} \end{aligned}$$

With these equations it is clear how the left hand side of Eq.(2.6) transforms. Let us assume that the potential V only depends on the lengths $|\mathbf{x}_i - \mathbf{x}_j|$ of the difference vectors (as is the case for gravity) then it is not difficult to see that

$$-\frac{\partial V}{\partial \mathbf{x}_i}(\mathbf{x}_1, \ldots, \mathbf{x}_N) = -R(t)\frac{\partial V}{\partial \mathbf{x}'_i}(\mathbf{x}'_1, \ldots, \mathbf{x}'_N)$$

We therefore find that

$$\begin{aligned} m_i\Big\{ \frac{d^2\mathbf{x}'_i}{dt^2} + 2R^{-1}(t)\frac{dR(t)}{dt}\frac{d\mathbf{x}'_i}{dt} + R^{-1}(t)\frac{d^2R(t)}{dt^2}\mathbf{x}'_i(t) + R^{-1}(t)\frac{d^2\mathbf{r}}{dt^2}\Big\} \\ = -\frac{\partial V}{\partial \mathbf{x}'_i}(\mathbf{x}'_1, \ldots, \mathbf{x}'_N) \end{aligned} \tag{2.7}$$

This equation has a much more complicated form than Eq.(2.6). It is clear that the new coordinate frame is not invariant anymore under the transformations (2.2)-(2.4) and therefore does not represent an inertial frame. The homogeneity in time is is violated as a consequence of the presence of the explicitly time-dependent rotation matrices $R(t)$ and the explicit time-dependence of the translation vector $\mathbf{r}(t)$. The homogeneity in space is violated by the third term on the left hand side of Eq.(2.7) which is linear in $\mathbf{x}'_j(t)$. Finally, isotropy is violated by the last term on the left hand side. The simple form Eq.(2.1) or Eq.(2.6) seems only to apply in the privileged inertial frames which, by definition, respect isotropy and homogeneity of space and time. In particular, in such frames we have that when $V = 0$, that

$$m_i \frac{d^2 \mathbf{x}_i}{dt^2} = 0$$

or

$$\mathbf{x}_i(t) = \mathbf{x}_{i,0} + \mathbf{v}_i t.$$

Therefore in such coordinate systems objects on which no forces act move along straight lines. Note that this is a more a feature of the presence of the second time derivative in Newton's law rather than the isotropy and homogeneity of space and time. For example, if we would have changed the second derivative on the right hand side of Eq.(2.1) to a third derivative then the equations of motion would still have been invariant under space and time translations as well as under rotations. Moreover, in that case the equations of motion would still be invariant under Galilean transformations (in fact, they would even be invariant under transformations to linearly accelerated frames). However, free motions are in that case not given by straight lines. We mention it here, since one often defines inertial frames as frames in which free motion is given by straight lines. However, the more fundamental property of inertial frames is the isotropy and homogeneity of space and time in such frames.
Let us finally address a particularly useful transformation between inertial frames. This is the transformation to the center-of-mass frame. Let us consider a set of N masses m_j which interact which each other (but not with any outside forces). Their motion is described with respect to some inertial frame by the equations

$$m_j \frac{d^2 \mathbf{x}_j}{dt^2} = \sum_{k \neq j}^{N} \mathbf{F}_{jk}(\mathbf{x}_j - \mathbf{x}_k) \tag{2.8}$$

where $\mathbf{F}_{jk}$ is the force on mass m_j by another mass m_k. Here we do not care about the precise nature of these forces. The masses may interact by gravitational forces, or be connected by springs or interact in some other way. We only assume that the forces are such that $\mathbf{F}_{jk} = -\mathbf{F}_{kj}$ and that the force $\mathbf{F}_{jk}$ only depends on the vector $\mathbf{x}_j - \mathbf{x}_k$ and respects the isotropy of space (i.e. transforms appropriately under the transformation Eq.(2.4)). Then we introduce the center-of-mass vector $\mathbf{X}$ by

$$\mathbf{X} = \frac{1}{M} \sum_{j}^{N} m_j \mathbf{x}_j \tag{2.9}$$

where $M = m_1 + \ldots + m_N$ is the sum of all masses. Then it follows that

$$M \frac{d^2 \mathbf{X}}{dt^2} = \sum_{j}^{N} m_j \frac{d^2 \mathbf{x}_j}{dt^2} = \sum_{j,k}^{N} \mathbf{F}_{jk} = 0 \tag{2.10}$$

since the last sum runs over all pairs of masses and $\mathbf{F}_{jk} = -\mathbf{F}_{kj}$. The center-of-mass $\mathbf{X}(t)$ therefore moves with constant velocity with respect to our inertial frame. The system that moves with the center of mass is therefore also an inertial frame. If we therefore define the new coordinates

$$\mathbf{x}'_j = \mathbf{x}_j - \mathbf{X} \tag{2.11}$$

then we are performing a Galilean transformation and Eq.(2.8) attains the same form

$$m_j \frac{d^2 \mathbf{x}'_j}{dt^2} = \sum_{k \neq j}^{N} \mathbf{F}_{jk}(\mathbf{x}'_j - \mathbf{x}'_k) \tag{2.12}$$

in the new coordinates. In the new system we have the useful relation

$$\sum_j^N m_j \mathbf{x}'_j = \sum_j^N m_j(\mathbf{x}_j - \mathbf{X}) = \sum_j^N m_j \mathbf{x}_j - M\mathbf{X} = 0 \tag{2.13}$$

This ends our introductory discussion of inertial frames. The concept of inertial frame introduces both a practical and a theoretical problem. Let us start with the practical problem in the next section. After that we discuss the more conceptual problem.

2.2 How to choose an inertial frame?

How do we choose an inertial frame in practice? The Earth is not an inertial frame since it rotates around its axis and is in accelerated motion with respect to the Sun. We find, however, experimentally that to a very good approximation any reference frame that moves at constant velocity with respect to the stars that can seen from Earth (and does not rotate with respect to them) forms an inertial frame. This is, in particular, true for the center-of-mass of the Solar System[1]. Indeed, in this reference frame the motions of the planets and other bodies in the Solar System are described to a very high accuracy by Eq.(2.1) The accurate description of celestial mechanics was one of the impressive successes of the Newton's theory of gravitation.
However, even accelerated reference frames may locally be regarded as inertial frames. Such inertial frames are formed by the frame of a test particle (which means that it has negligible mass) moving freely in a gravitational field. Let us, as an example, consider a small group of N asteroids orbiting the Sun (see Fig. 2.2) as described in the inertial frame attached to the center of mass of the Solar System.

Figure 2.2: A group of small masses orbiting the center of mass of the Solar System.

If we let the mass and the position of the Sun in this reference system be M and $\mathbf{X}$ and those of the asteroids (with much smaller masses) be m_j and $\mathbf{x}_j$ then the equation of motion of the asteroids is given by

$$m_j \frac{d^2 \mathbf{x}_j}{dt^2} = -GMm_j \frac{\mathbf{x}_j - \mathbf{X}}{|\mathbf{x}_j - \mathbf{X}|^3} - G\sum_{k \neq j}^N m_j m_k \frac{\mathbf{x}_j - \mathbf{x}_k}{|\mathbf{x}_j - \mathbf{x}_k|^3} \tag{2.14}$$

Let us now define the center of mass of the asteroids to be

$$\mathbf{R} = \frac{1}{\mu} \sum_{j=1}^N m_j \mathbf{x}_j \tag{2.15}$$

[1] Newton writes in his Principia that "The common center of gravity of the Earth, the Sun, and all planets is immovable." He referred to this center of mass as "the center of the system of the world" and assumed it to be at rest with respect to absolute space. This center of mass is located just inside or outside the surface of the Sun and its motion relative to the Sun is mainly determined by the positions of the large planets Jupiter and Saturn.

where $\mu = m_1 + \ldots + m_N$ is the total mass of the asteroids. This mass center satisfies the equation of motion

$$\mu \frac{d^2\mathbf{R}}{dt^2} = \sum_{j=1}^{N} m_j \frac{d^2\mathbf{x}_j}{dt^2} = -GM \sum_{j}^{N} m_j \frac{\mathbf{x}_j - \mathbf{X}}{|\mathbf{x}_j - \mathbf{X}|^3} \tag{2.16}$$

We now define the coordinates $\mathbf{x}'_j = \mathbf{x}_j - \mathbf{R}$ which describe the position of the asteroids with respect to their center of mass. Then in these coordinates we have

$$\mu \frac{d^2\mathbf{R}}{dt^2} = -GM \sum_{j}^{N} m_j \frac{\mathbf{x}'_j + \mathbf{R} - \mathbf{X}}{|\mathbf{x}'_j + \mathbf{R} - \mathbf{X}|^3} \tag{2.17}$$

$$m_j \left(\frac{d^2\mathbf{x}'_j}{dt^2} + \frac{d^2\mathbf{R}}{dt^2} \right) = -GMm_j \frac{\mathbf{x}'_j + \mathbf{R} - \mathbf{X}}{|\mathbf{x}'_j + \mathbf{R} - \mathbf{X}|^3} - G \sum_{k \neq j}^{N} m_j m_k \frac{\mathbf{x}'_j - \mathbf{x}'_k}{|\mathbf{x}'_j - \mathbf{x}'_k|^3} \tag{2.18}$$

Now we assume that

$$|\mathbf{x}'_j| \ll |\mathbf{R} - \mathbf{X}| \tag{2.19}$$

i.e. we assume that the asteroids in the small group are much closer to their common center of mass than to the Sun. In that case we can neglect $\mathbf{x}'_j$ compared to $\mathbf{R} - \mathbf{X}$ in the first terms on the right hand sides of Eqs. (2.17) and (2.18). We further have that $|\mathbf{X}| \ll |\mathbf{R}|$ due to the very large mass of the Sun such that we can write $\mathbf{R} \approx \mathbf{R} - \mathbf{X}$ and simplify the equations a bit further[2]. We these approximations we then obtain

$$\frac{d^2\mathbf{R}}{dt^2} = -GM \frac{\mathbf{R}}{|\mathbf{R}|^3} \tag{2.20}$$

$$m_j \frac{d^2\mathbf{x}'_j}{dt^2} = -G \sum_{k \neq j}^{N} m_j m_k \frac{\mathbf{x}'_j - \mathbf{x}'_k}{|\mathbf{x}'_j - \mathbf{x}'_k|^3} \tag{2.21}$$

We see that in the frame attached to the center of mass $\mathbf{R}(t)$ the asteroids follow the same gravitational law as in an inertial frame. This remains true as long as the asteroids do not wander to far from their common center of mass as we assumed that $|\mathbf{x}'_j| \ll |\mathbf{R}|$. So the inertial properties are only valid locally. We further see that the center of mass $\mathbf{R}(t)$ moves freely as a single particle in the gravity field of the Sun. Its motion is either elliptic, parabolic or hyperbolic, depending on the initial conditions imposed.

Another important example of a local inertial frame is the Earth-Moon system. This is just a special case of the previous derivation with only two small masses m_1 and m_2 representing the Moon and the Earth. In that case $\mathbf{R}(t)$ is given by the center of mass of the Moon-Earth system which describes an elliptic motion around the much more massive Sun. We may say that the Moon-Earth system is in a free fall around the Sun and systems that move freely in a gravitational field behave locally like an inertial frame. In fact, Einstein took this as a guiding principle to develop a relativistic theory of gravitation.

2.3 Conceptual difficulties

Let us now consider a conceptual difficulty with the concept of inertial frame. This is most easily illustrated with a simple example. We consider a system consisting of an observer and spring connected to two identical masses. We then consider two physical situations A and B :

[2]The asteroid belt is about 450 million kilometers removed from the Sun while the center of mass of the Solar System is on average one solar radius or about 0.7 million kilometers removed from the center of the Sun.

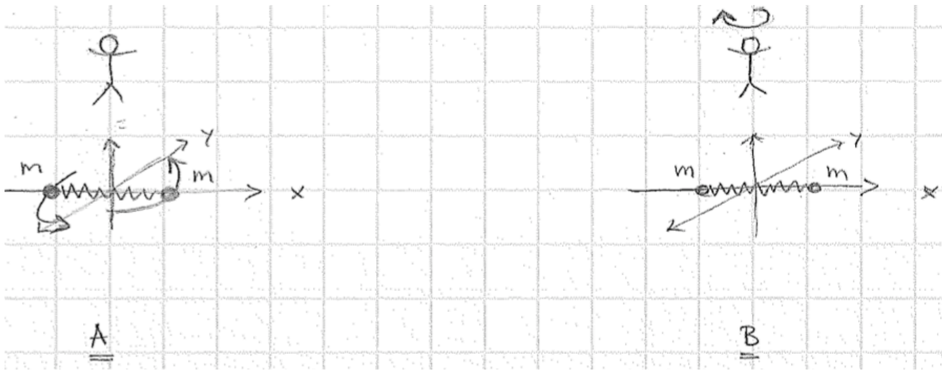

Figure 2.3: The physical situations of a rotating spring and resting observer A) and a rotating observer and a resting spring B)

A] In this case the spring is rotating with constant angular velocity with respect to the inertial frame at which the observer is at rest. The spring has a fixed length that does not change with time.

B] In this case the spring is at rest in an inertial frame and the observer is rotating with constant angular velocity with respect to this frame. Again the spring has a fixed length that does not change in time.

If one imagines a universe only containing the observer and the spring one would imagine that both cases are completely equivalent. Nevertheless, Newton's equations (2.6) predict that in the case A the spring is stretched out of its equilibrium length and will be longer than the spring in case B. It is instructive to see how Newton's equations achieve this result. Let us start with situation A. We use Eq.(2.6) in which the potential V is given by the harmonic potential

$$V(\mathbf{x}_1, \mathbf{x}_2) = \frac{k}{2}(|\mathbf{x}_1 - \mathbf{x}_2| - R_0)^2$$

where k is the spring constant and R_0 the equilibrium length of the spring. The equations (2.6) thus become

$$\begin{aligned} m\frac{d^2\mathbf{x}_1}{dt^2} &= -k(|\mathbf{x}_1 - \mathbf{x}_2| - R_0)\frac{\mathbf{x}_1 - \mathbf{x}_2}{|\mathbf{x}_1 - \mathbf{x}_2|} \\ m\frac{d^2\mathbf{x}_2}{dt^2} &= -k(|\mathbf{x}_1 - \mathbf{x}_2| - R_0)\frac{\mathbf{x}_2 - \mathbf{x}_1}{|\mathbf{x}_2 - \mathbf{x}_1|} \end{aligned}$$

Subtracting both equations then gives

$$\mu\frac{d^2\mathbf{x}}{dt^2} = -k(|\mathbf{x}| - R_0)\frac{\mathbf{x}}{|\mathbf{x}|} \tag{2.22}$$

where $\mathbf{x} = \mathbf{x}_1 - \mathbf{x}_2$ and $\mu = m/2$ is the reduced mass. A solution describing uniform rotation with fixed length $|\mathbf{x}| = R$ is given by

$$\mathbf{x}(t) = (R\cos(\Omega t), R\sin(\Omega t), 0)$$

where Ω is the angular velocity. Inserting this expression into Eq.(2.22) gives

$$-\mu\Omega^2 = -k\frac{R - R_0}{R} \quad \Rightarrow \quad R = \frac{R_0}{1 - \frac{\mu\Omega^2}{k}} > R_0 \tag{2.23}$$

when $\mu\Omega^2/k < 1$ [3]. We see that $R > R_0$ and therefore the rotating spring has increased its length as compared to its equilibrium length.
Let us now discuss case B where the spring is at rest in the inertial frame (and for instance $\mathbf{x} = (R_0, 0, 0)$ solves Eq. (2.22) in that case). In this case we apply Eq.(2.7) . The rotating frame with coordinates $\mathbf{x}'$ is related to the inertial frame with coordinates $\mathbf{x}$ by the rotation matrix

$$\mathbf{x} = \begin{pmatrix} \cos(\Omega t) & \sin(\Omega t) & 0 \\ -\sin(\Omega t) & \cos(\Omega t) & 0 \\ 0 & 0 & 1 \end{pmatrix} \mathbf{x}' = R(t)\mathbf{x}' \tag{2.24}$$

From this expression we can calculate that Eq.(2.7) attains the form

$$\mu\left(\frac{d^2\mathbf{x}_i'}{dt^2} - 2\Omega\,\mathbf{e}_z \times \frac{d\mathbf{x}_i'}{dt} - \Omega^2\mathbf{x}_i'\right) = -\frac{\partial V}{\partial \mathbf{x}_i'}(\mathbf{x}_1', \ldots, \mathbf{x}_N')$$

where $\mathbf{e}_z = (0, 0, 1)$. In particular for the vector $\mathbf{x}' = \mathbf{x}_1' - \mathbf{x}_2'$ we have

$$\mu\left(\frac{d^2\mathbf{x}'}{dt^2} - 2\Omega\,\mathbf{e}_z \times \frac{d\mathbf{x}'}{dt} - \Omega^2\mathbf{x}'\right) = -k(|\mathbf{x}'| - R_0)\frac{\mathbf{x}'}{|\mathbf{x}'|} \tag{2.25}$$

It is immediately clear from Eq.(2.24) that a solution to Eq.(2.25) is given by

$$\mathbf{x}'(t) = R^{-1}(t)\begin{pmatrix} R_0 \\ 0 \\ 0 \end{pmatrix} = R_0\begin{pmatrix} \cos(\Omega t) \\ \sin(\Omega t) \\ 0 \end{pmatrix}$$

for which $|\mathbf{x}'| = R_0$ and consequently the right hand side of Eq.(2.25) is zero. From the viewpoint of the rotating observer the stretching of the spring is prevented by the so-called "fictitious" forces in the second and third term on the left hand side of Eq.(2.25).
A further interesting case is when the observer is rotating together with the spring at the same constant angular velocity as seen from the inertial observer. In that case the spring is at rest in the rotating frame and consequently the time derivatives of $\mathbf{x}'$ vanish. Then Eq.(2.25) reduces to

$$-\mu\Omega^2\mathbf{x}' = -k(|\mathbf{x}'| - R_0)\frac{\mathbf{x}'}{|\mathbf{x}'|}$$

This equation has the solution (2.23) and the rotating observer finds that the spring is stretched by the fictitious force $\mathbf{F} = -\mu\Omega^2\mathbf{x}'$.
We see that rotation has an absolute character. In a rotating frame fictitious forces are present that lead to physically observable effects (such as the stretching of a the spring) whereas in frames at rest with respect to an inertial frame they are absent. This is conceptually difficult to understand. What distinguishes situations A and B in Fig 2.3? Newton was very much aware of this problem. He explained the difference between situation A and B by introducing the concept of *absolute space*. An object rotating or accelerating with respect to absolute space experiences fictitious forces whereas an object moving with constant velocity with respect to absolute space does not. So in case A the spring rotates with respect to the absolute space, and as a consequence gets stretched. In case B the observer rotates with respect to the absolute space and consequently feels fictitious forces whereas the spring is at rest with respect to the absolute space and does not experience any forces. The obvious question that arises from this viewpoint is, of course, how the spring in case A "knows" that it is rotating with respect to the absolute space. One would expect that obtaining this information requires an interaction

[3]When this condition is violated a solution with constant length of the spring is not possible. Of course, if the spring rotates too fast it will break and the harmonic approximation will also break down

between the spring and the absolute space. How this would happen is not at all explained within Newtonian mechanics. However, Newton needed absolute space to explain the problem above, and placed the fixed stars that we see from the Earth at rest in this absolute space and therefore rotation with respect to the fixed stars is a rotation with respect to absolute space. He further put the center of mass of the Solar System at rest with respect to absolute space by assumption.

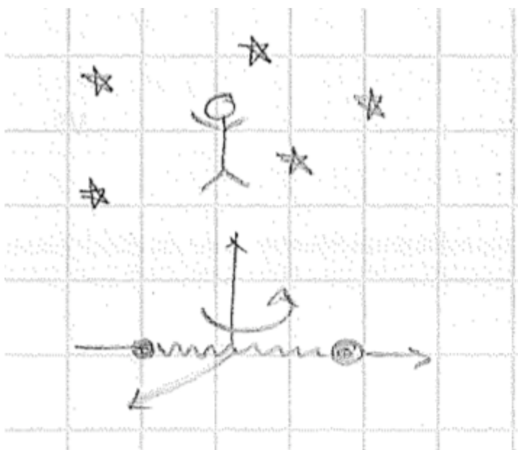

Figure 2.4: A spring rotating with respect to the distant stars

Newton did not propose a causal connection between the distant stars and the inertial forces. The suggestion was first made by Mach who noted that the distant stars seem to be the cause of the inertial forces in the rotating frame. The whole issue was finally clarified by Einstein who derived that rotation is always relative with respect to the local gravitation field. In Fig 2.4 the local gravitation field is defined by the distant mass distributions (and possibly by nearby planets such as the Earth). In Einstein's gravitation theory (the general theory of relativity) freely "falling" objects in a gravitational field define local inertial frames and rotations with respect to these frames are rotations with respect to the local gravitational field. A rotating object can interact with the gravitational field and even drag it along (this is known as the Lense-Thirring effect and is observed by satellites orbiting the Earth). It would go too far to discuss this interesting physics at this point. The main point of the discussion here was meant to give a clear discussion of the concept of inertial frame.

2.4 Rotating frames and general space-time coordinates

In the example of the previous section we discussed rotating frames within the context of Newtonian mechanics. The natural question that immediately arises is how we can describe rotating frames starting from a Lorentz invariant theory of inertial frames. We will investigate this question in this Section. Let us imagine a Lorentzian inertial frame O_1 that uses coordinates (t, x, y, z) . Let us in the origin O of the coordinate frame be a rotating disc with radius R that rotates with constant angular velocity ω around the z-axis with respect system O_1.

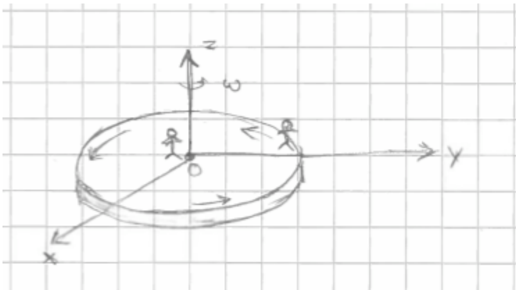

Figure 2.5: An observer on a disc rotating with constant angular velocity ω with respect to an inertial system.

Let us further imagine on the disc a specially marked material point P, such as a marker made with a pen on the disc. The point makes a rotation motion around the origin O. Let at $t = 0$ the position of the point P be given by the spatial coordinate $(x, y, z) = (r\cos\theta_0, r\sin\theta_0, 0)$. Then its motion with respect to O_1 is described by the curve

$$\mathbf{x}(t) = (r\cos(\theta_0 + \omega t), r\sin(\theta_0 + \omega t), 0) \tag{2.26}$$

and has with respect to O_1 the velocity

$$\mathbf{v}(t) = \frac{d\mathbf{x}}{dt} = (-\omega r\sin(\theta_0 + \omega t), \omega r\cos(\theta_0 + \omega t), 0)$$

We therefore have $v = |\mathbf{v}(t)| = \omega r$. Since no material point can move faster than the speed of light with respect to O_1 we must have that the radius R of the disc is restricted by $\omega R < c$. Let us now consider a clock that is at rest at point P of the rotating disc. From the viewpoint of O_1 this clock moves at velocity $v = \omega r$. This clock is instantaneously at rest in an inertial frame that moves at speed $v = \omega r$ with respect to O_1 and therefore we deduce that this clock will run slow by a factor $\sqrt{1 - v^2/c^2} = \sqrt{1 - \omega^2 r^2/c^2}$ as compared to a clock at rest in system O_1. So if dt is a small increment in proper time for a clock in the inertial frame then the corresponding increment in proper time $d\tau$ for a clock on the disc is given by

$$d\tau = \sqrt{1 - \omega^2 r^2/c^2}dt. \tag{2.27}$$

In general O_1 will therefore see that clocks at rest with respect to the disc run slower the further they are removed from the center of the disc. Let us now consider the spatial coordinates. Due to the symmetry of the problem it is natural to introduce spatial cylindrical coordinates (r, θ, z) related to the old spatial coordinates (x, y, z) by

$$\begin{aligned} x &= r\cos\theta \\ y &= r\sin\theta \\ z &= z \end{aligned} \tag{2.28}$$

We then consider two nearby points (r, θ_0) and $(r, \theta_0 + d\theta)$ on the disc at equal distance r from the origin but with an angular coordinate differing by a small amount $d\theta$. The distance between these points can be deduced in system O_1 by taking a snapshot of the disc at a given time t. Then Eq.(2.26) tells that the distance between the points is given by $dl = rd\theta$. Now consider a moving inertial system in which the two points are at rest. This can only be done approximately since the direction of the velocity of the two points is slightly different but this difference becomes negligible for small enough $d\theta$. In this frame the distance between the two points is dl'. Since the inertial system has velocity $v = \omega r$ with respect to O_1 we find that from the viewpoint of O_1 the distance dl' is Lorentz contracted to $rd\theta = dl = dl'\sqrt{1-\omega^2 r^2/c^2}$ and therefore

$$dl' = \frac{rd\theta}{\sqrt{1-\omega^2 r^2/c^2}}$$

If we, on the other hand, had considered two points that have the same angular coordinate θ but are separated by a radial coordinate dr then we would not observe any Lorentz contraction since the separation vector is orthogonal to the velocity of both points and in this case $dl = dl'$ and therefore both observers agree on the length of the radius of the disc. We see that in a frame moving with the disc the circumference of a circle with radial coordinate r is given by

$$l' = \int_0^{2\pi} \frac{rd\theta}{\sqrt{1-\omega^2 r^2/c^2}} = \frac{2\pi r}{\sqrt{1-\omega^2 r^2/c^2}} > 2\pi r$$

and therefore conclude that for the moving observer the coordinates r and θ do not play the role of Euclidean coordinates. So far we used the inertial system O_1 as a reference to reach our conclusions. The question then arises how the physical world looks like for another observer O_2 at rest with respect to the disc. The first problem that observer O_2 faces is how to define a coordinate system. The first thing to comes to mind is to define space and time coordinates with a direct physical meaning as we did in Eqs.(1.33) and (1.34) using Einstein synchronization. However, with respect to the rotating disc light rays move along curved paths and we immediately run into the problem how to synchronize distant coordinate clocks. After some reflection, however, we realize that the actual choice of coordinates is irrelevant, as any choice will do. The situation is similar to defining a coordinate system on a general curved surface, like the surface of the Earth, on which no simple Euclidean coordinate system is possible. This, however, does not prevent us from introducing a coordinate system on the Earth's surface. Any coordinate system that labels a point on the Earth's surface uniquely is equally valid. The only problem is how to relate these coordinates to coordinate independent quantities, such as the distance as measured by local observers on Earth between two cities with certain coordinates, but this is a solvable problem for any coordinate system that we like to choose. We can do the same for observer O_2, we can choose any space-time coordinate system we like. The only requirement is that it labels any space-time event uniquely. We will worry later about its relation to times and lengths measured by local clocks and measuring rods in the vicinity of the observer on the disc. Let us discuss this issue from a more general point of view. To shorten the notation we write denote the coordinates of the inertial frame O_1 by (y^0, y^1, y^2, y^3) (the use of super-indices will be explained in later Chapters). These are not necessarily Cartesian coordinates. They could, for instance in our example be the coordinates (t, x, y, z) or (t, r, θ, z). For the non-inertial frame O_2 we will use new coordinates (x^0, x^1, x^2, x^3). The particular choice of these coordinates is irrelevant as long as it labels each space-time event uniquely. They also have, in general, no physical meaning which implies that physically observable quantities should eventually be independent of them. The only thing we need is that there is a one-to-one relation between the coordinates y^j and x^j, i.e. there are invertible functions $y^j(x^0, x^1, x^2, x^3)$ relating them. Let us now start by considering the standard Minkowski coordinates for system O_1, i.e. $(y^0, y^1, y^2, y^3) = (t, x, y, z)$ Consider now the infinitesimal distance

$$ds^2 = -c^2 dt^2 + dx^2 + dy^2 + dz^2 = \sum_{\mu,\nu=0}^{3} g_{\mu\nu} dy^\mu dy^\nu \tag{2.29}$$

between two space-time points, where $g_{\mu\nu}$ is a diagonal matrix with diagonal elements $(-c^2, 1, 1, 1)$. We have seen in the previous Chapter that this expression is invariant under Lorentz transformations in the sense that a Lorentz transformation transforms this expression into another diagonal quadratic form with identical constant diagonal elements. Nothing, however, prevents us to transform to arbitrary other coordinates (x^0, x^1, x^2, x^3) using

$$dy^\mu = \sum_{\rho=0}^{3} \frac{\partial y^\mu}{\partial x^\rho} dx^\rho$$

This transforms Eq.(2.29) into

$$ds^2 = \sum_{\rho,\sigma=0}^{3} g'_{\rho\sigma} dx^\rho dx^\sigma \tag{2.30}$$

where

$$g'_{\rho\sigma} = \sum_{\mu,\nu=0}^{3} g_{\mu\nu} \frac{\partial y^\mu}{\partial x^\rho} \frac{\partial y^\nu}{\partial x^\sigma}$$

is a symmetric matrix (or more properly tensor, but this will be discussed in much more detail in the next Chapter). The expression (2.30) is now, in general, not diagonal anymore and nor are the coefficients $g'_{\rho\sigma}$ constant. However, the transformation obviously did not change the value of ds^2, it has a value that is the same in any coordinate system. A simple example is given by a simple transformation to cylindrical coordinates (t, r, θ, z) using Eq.(2.28) which gives

$$\begin{aligned} dx &= \cos\theta dr - r\sin\theta d\theta \\ dy &= \sin\theta dr + r\cos\theta d\theta \end{aligned}$$

and

$$ds^2 = -c^2dt^2 + dx^2 + dy^2 + dz^2 = -c^2dt^2 + dr^2 + r^2d\theta^2 + dz^2 \tag{2.31}$$

So far, nothing special has happened. Things become more interesting when transform to a moving frame. Let us consider the system O_1 with coordinates (t, r, θ, z) and define new coordinates $(x^0, x^1, x^2, x^3) = (t', r', \theta', z')$

$$\begin{aligned} t' &= t \\ r' &= r \\ \theta' &= \theta - \omega t \\ z' &= z \end{aligned} \tag{2.32}$$

This is a transformation to the frame O_2 rotating together with the disc since the points in the new frame that have are spatially constant values, i.e. $(r', \theta', z') = (c_1, c_2, c_1)$ for constants c_1, c_2, c_3, are rotating uniformly with angular frequency ω with respect to O_1, i.e. $r = c_1, \theta = c_2 + \omega t, z = c_3$. It remains to relate the new coordinates to distances measured on the disc and times displayed by local clocks on the disc. In general, the new coordinates do not need to have a physical meaning. However, in this particular case, the choice of the time coordinate $t' = t$, which we will call the coordinate time, has a possible physical realization. An observer on the disc can simply decide to use the time displayed visually by a clock at rest in O_1, for example at the center of the disc, to label time values independent of what his/her local clock may display[4]. An event happening at a distance r' from the center will in this way be assigned the time value read visually from the clock at the center minus a correction $\Delta t'$ that observer O_2 at r' makes for the time it takes for a light signal to travel from the center of the disc to the observer. This correction is simply calculated. Let observer O_1 send out a light signal from the origin at time

[4]We could say that the coordinate time defines a non-standard clock, whereas the local proper time is displayed by a standard or physical clock.

$t = 0$. For observer O_1 this light signal simply moves radially outward long a straight line given by the coordinates

$$r = c\,t$$
$$\theta = \theta_0$$

(for simplicity we take $z = z' = 0$). From the coordinate transformation (2.32) it follows that observer O_2 sees that the light moves along the path

$$r' = c\,t'$$
$$\theta' = \theta_0 - \omega t'$$

Therefore from the viewpoint of the observer on the disc the light is spiraling outward. Nevertheless, both observer O_1 and O_2 agree that the light reaches the radius r' at time $t = t' = r'/c$ [5]. Therefore, the correction $\Delta t'$ that the observer O_2 needs to subtract from the visually observed clock time is r'/c. In this way O_2 can use the coordinate time t' to label events. What will be the relation between this time and the time displayed by a local clock at radius r'? To determine this we consider again the invariant ds^2. Since $dt = dt', dr = dr', d\theta = d\theta' + \omega dt'$ and $dz = dz'$ we obtain

$$\begin{aligned} ds^2 &= -c^2 dt^2 + dr^2 + r^2 d\theta^2 + dz^2 \\ &= -c^2(1 - \frac{\omega^2 r'^2}{c^2})dt'^2 + dr'^2 + r'^2 d\theta'^2 + 2\omega r'^2 dt' d\theta' + dz'^2 \end{aligned} \tag{2.33}$$

Let us now consider two events that according to observer O_2 happen at the space time-points (t', r', θ', z') and $(t' + dt', r', \theta', z')$, i.e. at same spatial point such that $dr' = d\theta' = dz' = 0$ but with a coordinate time dt' apart. Then clearly

$$ds^2 = -c^2(1 - \frac{\omega^2 r'^2}{c^2})dt'^2$$

Let us further consider an inertial observer O_3 (moving with velocity $v = \omega r$ with respect to O_1) in which the spatial point (r', θ', z') is instantaneously at rest. In this inertial frame $ds^2 = -c^2 d\tau^2$ where τ is the proper time recorded by a clock at rest at point (r', θ', z') in O_3. Since observers O_2 and O_3 are at rest with respect to each other at the same spatial position they agree on the proper time increment $d\tau$ of their local clocks. We therefore find that the proper time $d\tau$ in system O_2 and the coordinate time increment dt' are related by

$$d\tau^2 = (1 - \frac{\omega^2 r'^2}{c^2})dt'^2$$

and by integration (since r' is constant) we find

$$\tau(t', r') - \tau(t'_0, r') = \sqrt{1 - \left(\frac{\omega r'}{c}\right)^2}\,(t' - t'_0) \tag{2.34}$$

We therefore find that the proper time τ recorded by a physical clock in system O_2 at radius $r = r'$ is a factor of $\sqrt{1 - \omega^2 r^2/c^2}$ smaller than the coordinate time t' at the same spatial point. We already came to this conclusion above. Note that, unlike the case of relatively moving inertial observers, the observers O_1 and O_2 agree with each other that the clock at rest in O_2 is running slower than the one at rest in O_1. This is because the two systems are no longer equivalent,

[5] Of course, O_1 must choose θ_0 correctly to reach observer O_2. You can check that the correct choice is $\theta_0 = \bar{\theta} + \omega r'/c$, where $\bar{\theta}$ is the angular coordinate θ of O_2 at $t = 0$.

while O_1 is an inertial observer the system O_2 is not. Note that in Eq.(2.34) we still need to make a choice for the value of $\tau(r', t_0)$. A simple choice is $\tau(r', 0) = 0$ which means that we set all local clocks to time zero when the coordinate time $t' = 0$. This is a simple example of synchronization using coordinate time. In practice this is easily achieved, an observer at r' simple sets his clock to zero when he sees that the clock in the center reads the time $t = -r'/c$. At any later moment every clock on the disc will then read the time

$$\tau(t', r') = t' \sqrt{1 - \left(\frac{\omega r'}{c}\right)^2} \tag{2.35}$$

The ratio between the proper times τ_1 and τ_2 of two clocks at $r' = r_1$ and $r' = r_2$ is then given by

$$\frac{\tau_1}{\tau_2} = \sqrt{\frac{1 - \left(\frac{\omega r_1}{c}\right)^2}{1 - \left(\frac{\omega r_2}{c}\right)^2}} \tag{2.36}$$

We see that the coordinate time t' has disappeared from the expression. The ratio above is an experimentally accessible quantity independent of the coordinate system used and does not depend on the coordinate system used. The variables r_1 and r_2 have a direct physical meaning as the distance to the origin measurable by standard measuring sticks. We will discuss below in more detail how O_2 can measure distances on the rotating disc, but before we do that we will give simple physical application of Eq.(2.34). Let an observer in the origin send out periodic signals with a time difference Δt and hence a frequency $\nu_0 = 1/\Delta t$. Then according to Eq.(2.34) the observer at distance $r' = r$ will receive these signals with a period $\Delta\tau(r)$ on his local clock and hence with a frequency ν_r given by

$$\nu_r = \frac{\nu_0}{\sqrt{1 - \left(\frac{\omega r}{c}\right)^2}}$$

The rotating observer finds the signal to be blue-shifted. Reciprocally, if the rotating observer would send out periodic signals the central observer would see them to be red-shifted. This phenomenon is also known as *gravitational redshift* and has been well-verified experimentally. The reason for this name in our context is that the rotating observer can interpret the force he feels on the rotating disc as the presence of a gravitational field [6]. The redshift is then caused by light trying to climb out of a gravitational well.
So far the discussion was focussed on local time measurements. The next question is what a local observer will find for distance measurements. Rather than working this out directly for the rotating disc we first work out this question in general terms such that the general structure of the equations become clear. After that we will again return to the rotating disc. Let us consider a (in general non-inertial) frame with coordinates $x = (x^0, x^1, x^2, x^3)$ which have no physical meaning apart from the fact that they label space-time points uniquely[7]. In the coordinates the invariant line element ds^2 attains the form

$$ds^2 = \sum_{\mu,\nu=0}^{3} g_{\mu\nu} dx^\mu dx^\nu$$

[6]This is Einstein's equivalence principle that forms the foundation of the theory of general relativity. The rotating disc played an important role Einstein's early work to understand gravity in terms of the deformation of space-time.

[7]Just like Matterhorn and Mont Blanc are labels without any physical meaning for two points on the surface of the Earth. However, the distance between these points does have a physical meaning, as well as the rate of local clocks at these points.

Let us first consider the case that $dx^\nu = 0$ for $\nu = 1, 2, 3$ while $dx^0 \neq 0$ i.e. we consider two space-time event at the same spatial position such that $ds^2 = g_{00}(x)(dx^0)^2$. By considering an inertial observer instantaneously at rest at point (x^1, x^2, x^3) we find, as in our example above, that $ds^2 = -c^2 d\tau^2$ where τ is the proper time of a clock at (x^1, x^2, x^3). We therefore find that

$$d\tau = \frac{1}{c}\sqrt{-g_{00}(x)}\, dx^0 \tag{2.37}$$

which is the generalization of the equation that we derived above for our example of the rotating disc. So we know how to relate coordinate time x^0 to local proper time. The next task is to discuss distances. We noted above that the measurement of distances with light signals in a general inertial frame is complicated by the fact that light moves along curved paths in a general frame. This problem vanishes for infinitesimally close points since in that case the curvature of the path can be neglected. Let us therefore consider a spatial point A with coordinates (x^1, x^2, x^2) and a neighboring point B with coordinates $(x^1 + dx^1, x^2 + dx^2, x^3 + dx^3)$. Now we send a light signal from B to A which is subsequently reflected in A and received back in B. The situation is depicted graphically in Fig.(2.6).

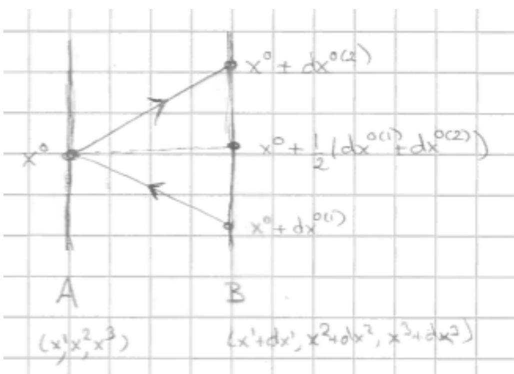

Figure 2.6: An observer in B sends out a light signal to a nearby observer A which is reflected back to A. The coordinate time which for observer B according Einstein synchronization is simultaneous with the reflection at coordinate time x^0 in A is indicated in the figure.

Given the spatial coordinates of A and B, what is now the coordinate time distance dx^0 for a light signal arriving at or leaving these points? This question can be answered by considering the invariant line element ds^2. Since we have $ds^2 = 0$ for a light signal the value dx^0 can be determined from the equation

$$0 = \sum_{\mu,\nu=0}^{3} g_{\mu\nu} dx^\mu dx^\nu = \alpha + 2\beta\, dx^0 + g_{00}(dx^0)^2 \tag{2.38}$$

where we defined

$$\alpha = \sum_{\mu,\nu=1}^{3} g_{\mu\nu} dx^\mu dx^\nu$$

$$\beta = \sum_{\nu=1}^{3} g_{0\nu} dx^\nu$$

The quadratic equation (2.38) is easily solved to give

$$dx^0 = -\frac{1}{g_{00}}\left(\beta \pm \sqrt{\beta^2 - \alpha\, g_{00}}\right) \tag{2.39}$$

There are two solutions, $dx^{0(1)}$ and $dx^{0(2)}$ to this equation, which is easy to understand physically (see Fig.(2.6)). One solution corresponds to the coordinate time $x^0 + dx^{0(1)}$ that B sends the light signal and another one $x^0 + dx^{0(2)}$ corresponds to the coordinate time at which B receives the signal. We therefore see that $dx^{0(1)} < 0$ while $dx^{0(2)} > 0$. Since $g_{00} < 0$ and $\alpha > 0$ (since it represents a spatial distance) we see that the square root term in Eq.(2.39) is larger than $|\beta|$ and we therefore have that $dx^{0(1)}$ and $dx^{0(2)}$ respectively correspond to the minus and plus sign in Eq.(2.39) [8]. The coordinate time passed between emission and reception of the light signal in B is then given by

$$dx^{0(2)} - dx^{0(1)} = -\frac{2}{g_{00}}\sqrt{\beta^2 - \alpha\, g_{00}} \tag{2.40}$$

The proper time interval $d\tau$ passed on a local clock in B between these two events is therefore according to Eq.(2.37) given by

$$d\tau = \frac{1}{c}\sqrt{-g_{00}(x)}\,(dx^{0(2)} - dx^{0(1)}) = \frac{2}{c}\frac{1}{\sqrt{-g_{00}}}\sqrt{\beta^2 - \alpha\, g_{00}} \tag{2.41}$$

We have calculated now calculated the return time of the light signal as measured in B. How can we use this do measure the distance to A? We simply use the same definition that we used in the case of inertial frames and *define* that the distance dl to point A is given by $dl = c\, d\tau/2$, i.e. for points further away it takes longer for the signal to return. We might call the distance determined in this way the *radar distance* between objects. By this operational definition of distance we find from Eq.(2.41) that

$$dl^2 = \alpha - \frac{\beta^2}{g_{00}} = \sum_{\mu,\nu=1}^{3}(g_{\mu\nu} - \frac{g_{0\mu}g_{0\nu}}{g_{00}})dx^\mu dx^\nu = \sum_{\mu,\nu=1}^{3}\tilde{g}_{\mu\nu}dx^\mu dx^\nu \tag{2.42}$$

where we defined the spatial three-dimensional metric

$$\tilde{g}_{\mu\nu} = g_{\mu\nu} - \frac{g_{0\mu}g_{0\nu}}{g_{00}} \qquad (\mu,\nu = 1,2,3) \tag{2.43}$$

With this new tensor we can measure the distance along any curve in the general frame. Let us illustrate this by going back to the example of the moving frame. The observer O_2 uses coordinates $(x^0, x^1, x^2, x^3) = (t', r', \theta', z')$ with the metric tensor $g_{\mu\nu}$ given in Eq.(2.33). The tensor $\tilde{g}_{\mu\nu}$ of Eq.(2.43) is then readily calculated to be given by

$$dl^2 = dr'^2 + \frac{r'^2}{1 - \frac{\omega^2 r'^2}{c^2}}d\theta'^2 + dz'^2 \tag{2.44}$$

This is the spatial metric that an observer on the rotating disc finds when he or she starts to carry out local experiments with light signals. It is clear that this metric is non-Euclidean (except of course when $\omega = 0$). The observer on the rotating disc, for instance, finds that a radial line (i.e. $d\theta' = dz' = 0$) from the center of the disc to point $r' = R$ has the length

$$l = \int_0^R dr' = R$$

while the circle with radius $r' = R$ has the circumference

$$l = \int_0^{2\pi}\frac{Rd\theta}{\sqrt{1 - \frac{\omega^2 R^2}{c^2}}} = \frac{2\pi R}{\sqrt{1 - \frac{\omega^2 R^2}{c^2}}}$$

[8] Note that the sign of β is arbitrary. For instance, if we let our disc rotate in the opposite direction by replacing ω by $-\omega$ the sign of β changes.

Therefore the ratio between the circumference and the radius is not 2π as in Euclidean geometry but larger (which is what mathematicians call hyperbolic geometry). We can also calculate that, with the exception of radial lines, the shortest distance between two point on the disc (a so-called geodesic curve) is no longer a straight line as in Euclidean geometry. It is clear that these distances and curves are independent of the coordinate system used in the reference system O_2. At this point it is useful to make a distinction between reference frames and coordinate transformations. The observer O_2 may instead of the coordinates $(x^0, x^1, x^2, x^3) = (t', r', \theta', z')$ have used other internal coordinates y^μ related to the old ones by

$$\begin{aligned} y^0 &= y^0(x^0, x^1, x^2, x^3) \\ y^j &= y^j(x^1, x^2, x^3) \quad (j = 1, 2, 3) \end{aligned} \tag{2.45}$$

where the new spatial coordinates y^j for $j = 1, 2, 3$ are only functions of the old spatial coordinates and not of x^0 while the new time coordinate y^0 is allowed to depend on all old coordinates [9]. In this way the new spatial coordinates of a point at rest on the disc will stay time-independent as it should for a co-moving observers. We can thus say that a reference frame is an equivalence class of transformations of the form of Eq.(2.45). The spatial metric $\tilde{g}_{\mu\nu}$ is invariant (or more precisely transforms as a tensor) under such restricted transformations (You can check this as an exercise for yourself). More general coordinate transformations, such as (2.32) transform between reference frames as well.

Now that we have defined local clock times and local distance measurements the next question is how an observer on the disc uses these times and distance to calculate velocities. For the measurement of velocity we run into the problem that we need two nearby positions and two times to define it, but now in our reference system clocks at two different positions run at a different rate so we need to be careful in how to calculate the time difference. This problem can not be solved without agreeing on a definition of simultaneity of two nearby events in our reference frame, which requires a synchronization of clocks. We did not have to bother about this issue before since we did not need to compare clocks at different positions as the metric tensor $\tilde{g}_{\mu\nu}$ was deduced using the clock times of one and same clock in point B only. The synchronization of clocks can be achieved in different ways. We can, for instance, use the Einstein synchronization that we discussed extensively in Section 1.3. Following Eq.(1.33) of that Section we define nearby clocks at points A and B to be synchronized when the coordinate time of reflection x^0 in A corresponds to the time

$$x_B^0 = x^0 + \frac{1}{2}(dx^{0(1)} + dx^{0(2)}) = x^0 - \frac{\beta}{g_{00}} \tag{2.46}$$

for observer B. The coordinate clock times that are simultaneous by this definition are marked in Fig.(2.6). In other words, two event events differing in coordinate time $dx^0 = x_B^0 - x^0$ and spatial coordinates $dx^\nu (\nu = 1, 2, 3)$ are defined to be simultaneous when

$$dx^0 = -\frac{1}{g_{00}} \sum_{\nu=1}^{3} g_{0\nu} dx^\nu \tag{2.47}$$

where we used the explicit form of β. From Eq.(2.46) it now follows that (see also Fig.(2.6)) according to an observer in B the coordinate time $d\tilde{x}^0$ it takes for the reflected light signal to travel from A to B is given by

$$d\tilde{x}^0 = x^0 + dx^{0(2)} - (x^0 + \frac{1}{2}(dx^{0(1)} + dx^{0(2)})) = dx^{0(2)} + \frac{\beta}{g_{00}} \tag{2.48}$$

[9] One usually also demands that $\partial y^0 / \partial x^0 > 0$ in order not to reverse the orientation of time.

which according to Eq.(2.37) amounts to a proper time interval $d\tilde{\tau}$ of

$$d\tilde{\tau} = \frac{1}{c}\sqrt{-g_{00}}(dx^{0(2)} + \frac{\beta}{g_{00}}) \tag{2.49}$$

The velocity of the reflected signal is then simply given by $dl/d\tilde{\tau}$. Let us see what this gives. The expression $d\tilde{\tau}$ can used to write Eq.(2.38) as

$$0 = -c^2 d\tilde{\tau}^2 + \alpha - \frac{\beta^2}{g_{00}} = -c^2 d\tilde{\tau}^2 + dl^2 \tag{2.50}$$

where we used Eq.(2.42). Therefore the velocity of the light signal according to an observer B is

$$\frac{dl}{d\tilde{\tau}} = c$$

which at first sight may not be very surprising. However, it is important to realize that this result is a consequence of the definition Eq.(2.46) of simultaneity within our reference frame. If we use another synchronization we find that the velocity of light is not equal c and depends on the direction of the light signal. One has therefore to be careful in defining velocity in non-inertial frames.

Looking at our derivation we realize that the particular linear combination of the space and time increments on the right hand side of Eq.(2.49) formally allows us to get rid of the mixed space and time differentials in the metric. More precisely, we can write

$$\begin{aligned} ds^2 &= \sum_{\mu,\nu=0}^{3} g_{\mu\nu} dx^\mu dx^\nu \\ &= -c^2 \left(\frac{\sqrt{-g_{00}}}{c}(dx^0 + \frac{1}{g_{00}} \sum_{\nu=1}^{3} g_{0\nu} dx^\nu) \right)^2 + \sum_{\mu,\nu=1}^{3} \tilde{g}_{\mu\nu} dx^\mu dx^\nu \\ &= -c^2 d\tilde{t}^2 + dl^2 \end{aligned} \tag{2.51}$$

where we define the differential

$$d\tilde{t} = \frac{\sqrt{-g_{00}}}{c}(dx^0 + \frac{1}{g_{00}} \sum_{\nu=1}^{3} g_{0\nu} dx^\nu) \tag{2.52}$$

The interesting fact about the form of the metric in Eq.(2.51) is that it splits the line element ds^2 in two parts with a physical interpretation. The second part dl^2 represents the spatial metric as measured locally with the radar reflection method, whereas the condition $d\tilde{t} = 0$ guarantees that two nearby space-time points x^ν and $x^\nu + dx^\nu (\nu = 0, 1, 2, 3)$ are simultaneous (see Eq.(2.47)) according to Einstein synchronization. Furthermore when the spatial displacements are zero ($dx^\nu = 0$ for $\nu = 1, 2, 3$) then $d\tilde{t}$ coincides with the local proper time $d\tau$ recorded on a standard clock. We can therefore assign a useful physical meaning to $d\tilde{t}$. However, the differential $d\tilde{t}$ also has has an important deficiency, it is usually not a total differential. This means that Eq.(2.52) can, in general, not be integrated to a time variable $\tilde{t}$. If it would, then it would have the desirable property that $\tilde{t}(x^0, x^1, x^2, x^2) = K$, with K a constant, would represent an Einstein-synchronized surface of space-time points. In other words, by introducing $\tilde{t}$ as a new time variable two space-time points with the same value of $\tilde{t}$ would represent simultaneous events by Einstein synchronization. Let us, however, look at an interesting case where $d\tilde{t}$ can be integrated and gives a nice new insight in our familiar Lorentz transformation. Afterwards we return to the rotating disc where it can not be integrated.

We noted in our discussion of the rotating disc that the rotating observer could use any coordinates that he or she pleases. If this is the case for non-inertial observers this should, of course, also be the case for inertial observers. What is then so special about the Lorentz transformation? Can we then also use the Galilean transformation in the special theory of relativity? The short answer to these questions is that we are perfectly allowed to use the Galilean transformation, but that the coordinates used in the Lorentz transformation have a preferred physical interpretation. Let us work this out in more detail. We consider again an observer A (Alice) in an inertial with space-time coordinates (t, x, y, z) with the standard interpretation as discussed in Section 1.3. In particular the invariant line element of Alice has the form

$$ds^2 = -c^2dt^2 + dx^2 + dy^2 + dz^2 \tag{2.53}$$

Another observer B (Bob) moves with a velocity v with respect to Alice and used coordinates (t', x', y', z') related to the coordinates of Alice by the Galilean transformation

$$\begin{aligned} t' &= t \\ x' &= x - vt \\ y' &= y \\ z' &= z \end{aligned} \tag{2.54}$$

This is a coordinate transformation to a moving frame since the points with constant spatial coordinates in Bob's system are moving with velocity v with respect to Alice. In Bob's coordinates the invariant line element attains the form

$$ds^2 = -c^2(1 - \frac{v^2}{c^2})dt'^2 + 2vdx'dt' + dx'^2 + dy'^2 + dz'^2 \tag{2.55}$$

We now read off the metric tensor in the new coordinates and use Eq.(2.52) and (2.43) to calculate

$$\begin{aligned} d\tilde{t} &= \sqrt{1 - \frac{v^2}{c^2}} \left(dt - \frac{vdx'}{c^2(1 - \frac{v^2}{c^2})} \right) \\ dl^2 &= \frac{dx'^2}{1 - \frac{v^2}{c^2}} + dy'^2 + dz'^2 \end{aligned} \tag{2.56}$$

We see that in this case we can integrate $d\tilde{t}$ to obtain

$$\tilde{t} = t\sqrt{1 - \frac{v^2}{c^2}} - \frac{vx'}{c^2\sqrt{1 - \frac{v^2}{c^2}}} \tag{2.57}$$

where we choose the integration constant such that $\tilde{t} = 0$ when $x' = t' = 0$. Two space-time events with the same $\tilde{t}$ variable are now Einstein-synchronized and moreover $\tilde{t}$ records the proper time by a clock at rest in Bob's reference frame. We further see from Eq.(2.56) that Bob's spatial metric can be made into a standard Euclidean form by defining a new variable

$$\tilde{x} = \frac{x'}{\sqrt{1 - \frac{v^2}{c^2}}} \tag{2.58}$$

as well as $\tilde{y} = y$ and $\tilde{z} = z$, such that

$$dl^2 = d\tilde{x}^2 + d\tilde{y}^2 + d\tilde{z}^2$$

Bob can then use standard measuring sticks to map out his spatial continuum $(\tilde{x}, \tilde{y}, \tilde{z})$. How are Bob's new physically motivated coordinates now related to Alice's coordinates? Not surprisingly we find after a short calculation that

$$\tilde{x} = \frac{x - vt}{\sqrt{1 - \frac{v^2}{c^2}}}$$
$$\tilde{t} = \frac{t - \frac{vx}{c^2}}{\sqrt{1 - \frac{v^2}{c^2}}}$$

as well as $\tilde{y} = y$ and $\tilde{z} = z$. We have recovered the Lorentz transformation by defining new coordinates with a physical meaning. In general we can say that special relativity allows for general coordinate transformations but in the case of transformations between inertial frames the Lorentz transformations are preferred for reasons of physical interpretation. It is, therefore, in these coordinates that the physical laws (such as Maxwell's equations) attain their most transparent form. These facts are usually not know or told to students until they study general relativity. For this reason we have presented the example of the rotating disc in this Section. It gives the physics and the insights of general relativity without the need to bother about the laws of gravity at this point[10].

So how does this now work out for our rotating disc? In that case see from Eq.(2.52) and the explicit form of the coefficients $g_{\mu\nu}$ that

$$d\tilde{t} = \sqrt{1 - \frac{\omega^2 r'^2}{c^2}} \left(dt - \frac{\omega r'^2 d\theta'}{c^2 (1 - \frac{\omega^2 r'^2}{c^2})} \right) \tag{2.59}$$

Due to the dependence on r' this can not be integrated. This implies that the definition of simultaneity depends on the direction in which we decide to synchronize the clocks. If we, for instance, decide to synchronize clocks along the circle r' =constant, such that $dr' = 0$ then we can formally integrate Eq.(2.59) to

$$\tilde{t} = t \sqrt{1 - \frac{\omega^2 r'^2}{c^2}} - \frac{\omega r'^2}{c^2 \sqrt{1 - \frac{\omega^2 r'^2}{c^2}}} \theta' \tag{2.60}$$

The space-time points that satisfy $\tilde{t} = K$ with K represent simultaneous events according to Einstein synchronization. However, we now note something peculiar about this expression. If we let the angular variable run from 0 to 2π and make a full circle the variable $\tilde{t}$ does not return to its original value. This illustrates the fact that synchronization depends on the path chosen to synchronize and is mathematically caused by the fact that $d\tilde{t}$ is not a total differential. We can therefore not globally synchronize the clocks with the radar method. The situation is illustrated geometrically in the figure below where we plot the surface $\tilde{t} = 0$ for the case that we start synchronizing in the positive θ' direction starting from $\theta' = 0$.

[10]From the viewpoint of general relativity the only thing that is special about special relativity is that in this theory we can always find a coordinate transformation that transforms the metric to a global Minkowskian form (such as from the rotating observer on the disc back to the inertial one). This property disappears in the presence true gravity fields.

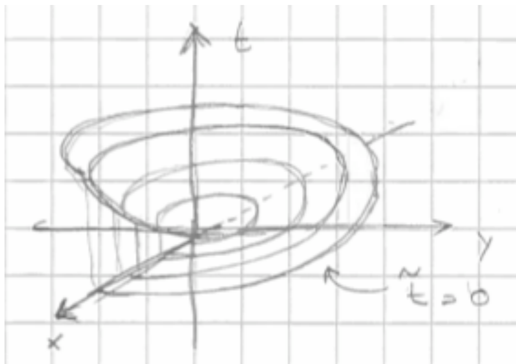

Figure 2.7: The surface $\tilde{t} = 0$ for synchronization in the positive θ' direction starting from $\theta' = 0$.

What about the radial direction? We see immediately from Eq.(2.59) that for $d\theta' = 0$ we have

$$d\tilde{t} = \sqrt{1 - \frac{\omega^2 r'^2}{c^2}} dt' \tag{2.61}$$

This equation can not be integrated, and we can not even define the variable $\tilde{t}$ locally. However, $d\tilde{t} = 0$ is equivalent to $dt' = 0$ and therefore the easiest thing we can do is use the coordinate time t' to define simultaneity for different radial positions. Our difficulties in defining simultaneity show that the concept does not have a well-defined experimentally accessible physical meaning, and that it is more question of definition. The problem is illustrated with the following example illustrated in Fig. 2.8.

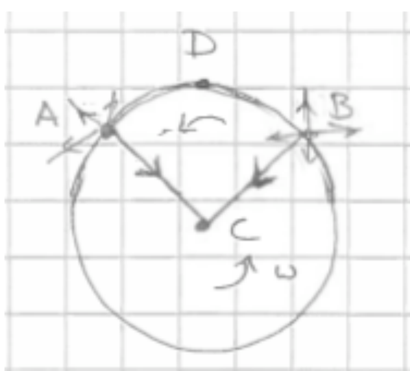

Figure 2.8: The rotating disc with Alice and Bob sending light flashes to Charlotte and Dilbert.

Two observers, Alice and Bob, at two different positions at the edge of a rotating disc send out a light flash in all directions. A third observer C (Charlotte) at the center of the disc receives the signals simultaneously at her clock, and since the distances AC and BC are equal she concludes the light flashes were sent at the same time. This means that Charlotte uses the central light flash method to define simultaneity which we discussed below Eq.(2.32). Observer D (Dilbert) who is located between Alice and Bob on the disc can come to two different conclusions depending on his definition of simultaneity. If he uses the coordinate time t' to define simultaneity he would agree that the flashes from Alice and Bob were simultaneous. For, instance, if Charlotte sends out a light flash in all directions to all observers on the edge of the disc then by the central-time synchronization the light arrives at the edge simultaneously for all observers. If this light subsequently gets reflected at the edge of the disc back to Charlotte all would agree that the light reflected by all observers arrives back in the center simultaneously. Dilbert then agrees on the simultaneity of the light reflections by Alice and Bob. If, Dilbert, however decides to use Einstein synchronization along the edge of the disc he would conclude on the basis of Eq.(2.7), since $\theta'_{\rm A} > \theta'_{\rm B}$, that $\tilde{t}_{\rm A} < \tilde{t}_{\rm B}$ and therefore that Alice send her light flash before Bob. This can be understood physically from the fact that Dilbert moves towards the light send out by Alice and away from the light send out by Bob (see also Fig.1.11 for a comparable physical situation). We therefore see that for a single observer in a non-inertial frame, Dilbert in our case, there is no preferred definition of simultaneity. This is a large break from the classical Newtonian thinking about time. Since we live on planet Earth, which is not an inertial frame, we can not say that something is happening on planet Mars "now". We first have to define what we mean by this. This concludes our discussion of the rotating disc. More illuminating discussions on the space-time physics of the rotating disc can be found in references [4, 5].

2.5 Physical meaning of particle trajectories

In the discussion of inertial frames we noted the invariance of Newton's equations under Galilean transformations. Similarly in the special theory of relativity we demand invariance of the physical laws under Lorentz transformations. In the general theory of relativity even invariance of physical laws under arbitrary coordinate transformations is imposed. Rather than performing endless coordinate transformations it is much more effective and illuminating to investigate what the invariant object in question actually is and to find a formulation of physical laws where coordinates do not even appear. This is exactly what we will start investigating next.
To get an intuitive insight into the concept of coordinate invariance we start with the example of a particle trajectory. We consider two observers moving with respect to each other. Each of them describes the trajectory of the same object in their own coordinate system. Observer A (Alice) is riding in a train at constant velocity with respect to the railroad and drops a stone out of the window. From the viewpoint of Alice the stone falls along a straight line until it hits the railroad. From the viewpoint of observer B (Bob) who is at rest with respect to the railroad the stone moves along a curved trajectory.

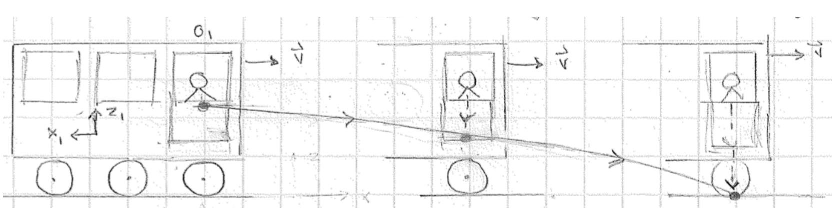

Figure 2.9: Dropping a stone from a moving train

Within Newtonian mechanics this trajectory would be a parabola (and within special relativity almost one with usual train speeds). This then immediately raises the question what is the "real" trajectory of the stone; is it a straight line or a parabola? The answer, of course, is that the straight line and the parabola are simply registrations of positions of the stone in arbitrarily chosen coordinate systems and do not have any absolute meaning.
However, there is something *absolute* we can say about the trajectory. Both observers completely agree on the position of the stone in relation to other material points. For instance, initially the stone was in the hand of Alice. At some point later the position of the stone was the same as the bottom of the train door (see Fig. 2.9) and even later the position of the stone is at the same position as where the train wheel touches the railroad track (see Fig. 2.9). We will call these

coincidences of material points *events*. Both Alice and Bob agree on the events. Moreover they agree on the value of the proper time τ of the stone at these coincidences of material points. With these observations we can give an absolute definition of the notion particle trajectory.
We start by defining the *space-time manifold* M as the collection of events (coincidences of material points). A particle trajectory is then defined as a curve γ connecting events in a continuous way. This curve can be parametrized by the proper time τ. A coordinate system on the space-time manifold is then defined to be a mapping φ from the set of events to the set of points $(t, x, y, x) \in \mathbb{R}^4$. This is displayed in the following figure:

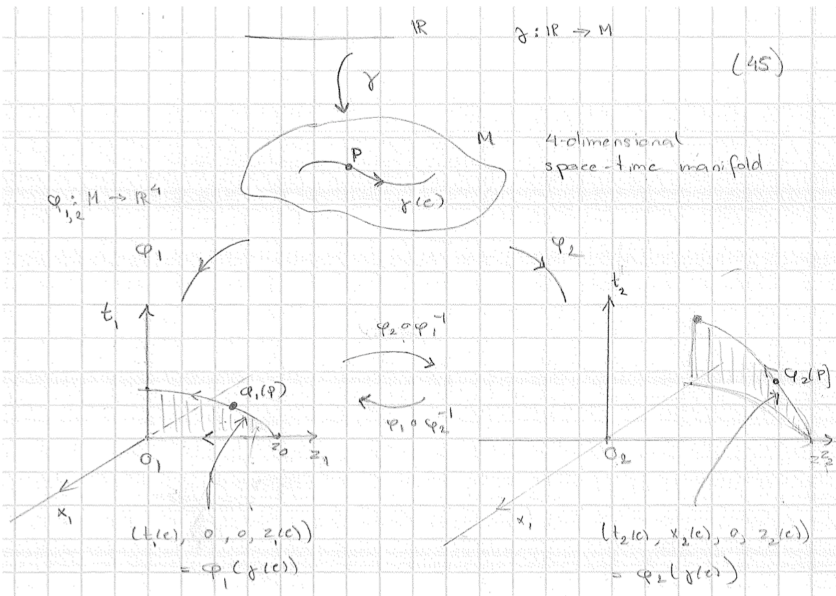

Figure 2.10: Different coordinate descriptions of the same curve γ connecting space-time events in a continuous way.

We have displayed the concept for the case of the stone falling from the train (displaying only the x, z-axes and leaving out the y-axis). Alice, who is traveling with the train, describes the trajectory γ of the stone in coordinate system φ_1 by

$$\varphi_1(\gamma(\tau)) = (t_1(\tau), 0, 0, z_1(\tau))$$

whereas Bob, who is at rest with respect to the railroad track, describes the trajectory γ of the stone in coordinate system φ_2 as

$$\varphi_2(\gamma(\tau)) = (t_2(\tau), x_2(\tau), 0, z_2(\tau))$$

The coordinate systems φ_1 and φ_2 are related by

$$\varphi_2(\gamma(\tau)) = (\varphi_2 \circ \varphi_1^{-1})(\varphi_1(\gamma(\tau)))$$

The mapping $\varphi_2 \circ \varphi_1^{-1}$ is therefore identical to the Lorentz transformation, i.e.

$$\varphi_2(\gamma) = \begin{pmatrix} t_2(\tau) \\ x_2(\tau) \\ 0 \\ z_2(\tau) \end{pmatrix} = \begin{pmatrix} \alpha & -\alpha\, v/c^2 & 0 & 0 \\ -\alpha\, v & \alpha & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} t_1(\tau) \\ 0 \\ 0 \\ z_1(\tau) \end{pmatrix} = \begin{pmatrix} \alpha\, t_1(\tau) \\ -\alpha\, v\, t_1(\tau) \\ 0 \\ z_1(\tau) \end{pmatrix}$$

where we denoted $\alpha = (1 - v^2/c^2)^{-1/2}$. In this case the mapping $\varphi_2 \circ \varphi_1^{-1}$ is a linear transformation. This is, however, a consequence of the fact that Alice and Bob have chosen a standard Euclidean coordinate system. If they, for instance, had chosen spherical coordinates instead then the transformation would have been non-linear and continuous mapping (which in physical applications is nearly always differentiable as well).
The concept of a space-time manifold becomes essential in the general theory of relativity. Clocks run at different rates at different locations and also spatial metric becomes non-Euclidian (we have seen clearly these to two features in the case of the rotating disc in the previous Section). This means that the physical lengths and times have no longer a direct relation with coordinates. The space-time coordinate looses its physical meaning and just becomes a label for an event. The only things that have physical meaning are the space-time coincidences of material points. With our example of the falling stone we have thus arrived in a natural way at the concept of a differentiable manifold. Let us therefore start again with the mathematical definition of a manifold and see what coordinate independent objects we can define on a manifold. This will lead us to the geometric concepts of vectors, tensor and geodesic curves. These objects will be then the building blocks of the physical laws.

2.6 Manifolds and coordinate maps

A manifold M is a set of elements on which differentiable functions are defined. A coordinate map or *chart* φ_i is a one-to-one map from a subset U_i of M to $\mathbb{R}^n$. The set of coordinate maps is the required to have the properties

1. The subsets U_i completely cover M, which means that every point in M belongs to at least one U_i
2. The mappings $\varphi_i \circ \varphi_j^{-1}$ are differentiable functions on the common domain $U_i \cap U_j$ on which they are defined

Pictorially we thus have

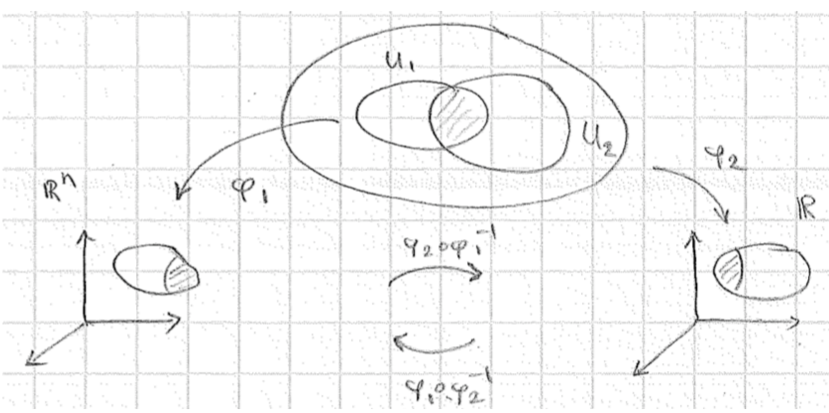

Figure 2.11: Different coordinate descriptions of the same physical event

This figure is essentially identical to fig 2.10. The main difference is that we did not require the coordinate maps φ_i to be defined on all of M. This has a practical reason. Often it is not possible to define a global coordinate map on all of M. A good example is the Earth's surface projected on a plane. For example, to study the North Pole and the South Pole areas we use two different maps. If the maps φ_i map one-to-one to a subset of $\mathbb{R}^n$ the set M is called an

n-dimensional manifold. The space-time manifold of figure 2.10 is a 4-dimensional manifold. Common examples of manifolds are two-dimensional surfaces. For instance, the surface of a sphere is defined by the set of points

$$M = \{(x, y, z) | x^2 + y^2 + z^2 = 1\}$$

Note that M is just defined by these triples of real numbers, there is nothing "outside" M although we often imagine the sphere to be embedded in a three-dimensional space. In fact, a main task of differential geometry is to describe the intrinsic properties of manifolds which are independent on whether they can be embedded in $\mathbb{R}^n$ for some value of n. This is similar to the way we read road maps. We need not think of anything outside the surface of the Earth to measure distances between various geographical locations using our maps. For the sphere (or the surface of the Earth) we can, for instance, have the coordinate maps

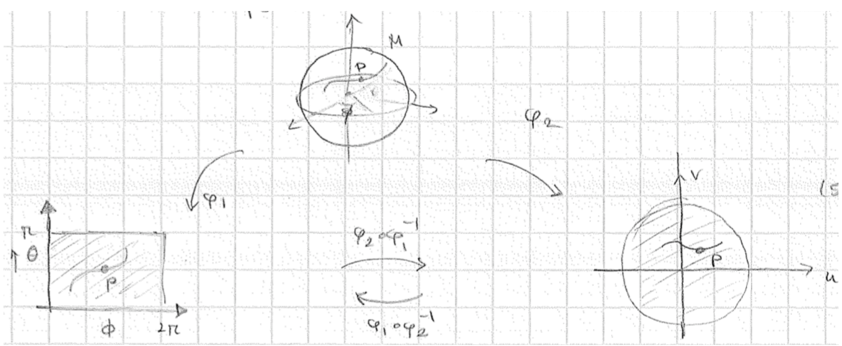

Figure 2.12: Different coordinate descriptions of points on the sphere (or, more physically, for objects on the surface of the Earth)

with

$$\begin{aligned} &\varphi_1(\cos\phi \sin\theta, \sin\phi \sin\theta, \cos\theta) = (\phi, \theta) \\ &\varphi_2(u, v, \sqrt{1 - u^2 - v^2}) = (u, v) \qquad u^2 + v^2 \leq 1 \end{aligned} \tag{2.62}$$

The first map is a familiar one that assigns to every point on the sphere a longitude ϕ and a latitude θ. It maps to a unique pair (ϕ, θ) except for the North and South pole of the sphere where the longitude ϕ is not uniquely defined. The map φ_2 maps every point in the northern hemisphere to a unique (u, v)-coordinate. The two coordinate systems are related by the mappings

$$\begin{aligned} (u, v) &= (\varphi_2 \circ \varphi_1^{-1})(\phi, \theta) = (\cos\phi \sin\theta, \sin\phi \sin\theta) \\ (\phi, \theta) &= (\varphi_1 \circ \varphi_2^{-1})(u, v) = (\arctan \frac{v}{u}, \arccos \sqrt{1 - u^2 - v^2}) \end{aligned}$$

These transformations convert coordinates of points on one map to coordinates on another map. Finally we discuss the mappings between manifolds.

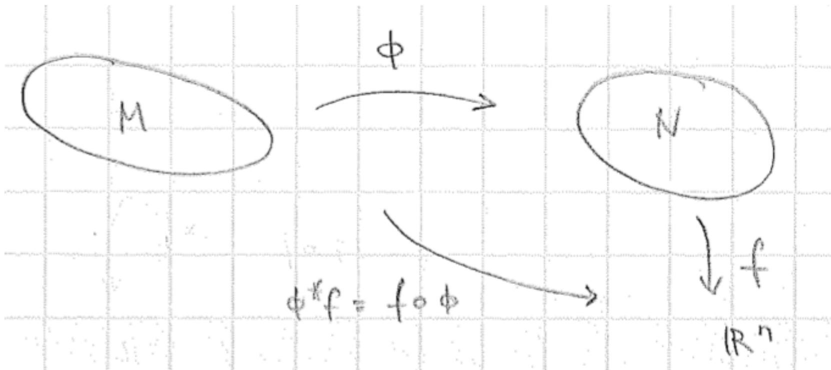

Figure 2.13: A mapping between two manifolds

If $f : N \to \mathbb{R}^n$ is mapping from a manifold N to $\mathbb{R}^n$ and $\phi : M \to N$ is a mapping from manifold M to N then we define

$$\phi^* f = f \circ \phi \tag{2.63}$$

The map $\phi^* f$ maps a point in M to $\mathbb{R}^n$ and is called the *pullback* of the map f by ϕ. A common example in which we need this mapping is when we want to embed a given manifold into another one. Let us give a simple example.

Imagine a region of space where the temperature in each point p is given by a function $T(p)$. One could think, for instance, that the temperature field describes a cloud C of some gas (which will be our manifold) in a distant solar system. The temperature field is then a mapping $T : C \to \mathbb{R}$ from the cloud to the real numbers. We can simply use a Cartesian coordinate system (x, y, z) to give coordinates to positions in the cloud. The manifold C is therefore taken to be equal to $\mathbb{R}^3$ and the coordination is simply the identity map[11]. Through this cloud moves a planet at relative velocity $\mathbf{v} = (v_\mathrm{x}, v_\mathrm{y}, v_\mathrm{z})$ with respect to the cloud, such that the position as a function of the time t of the planet's center is given by $\mathbf{x} = \mathbf{v}t$. The surface P of the planet will be our other manifold. We want to calculate the temperature at any point on the planet's surface. We do this by a pullback from C to P of the temperature field $T(x, y, z)$ defined on C . The surface of the sphere (which has radius one in appropriate units) is described by spherical coordinates as in Eq.(2.62). The time-dependent map $\phi_t : P \to C$ which maps a point of the surface P to a location in the cloud C is given by

$$\phi_t(\cos\phi\sin\theta, \sin\phi\sin\theta, \cos\theta) = (\cos\phi\sin\theta + v_\mathrm{x}t,\ \sin\phi\sin\theta + v_\mathrm{y}t,\ \cos\theta + v_\mathrm{z}t)$$

Then the pullback

$$(\phi_t^* T)(\cos\phi\sin\theta, \sin\phi\sin\theta, \cos\theta) = T(\cos\phi\sin\theta + v_\mathrm{x}t,\ \sin\phi\sin\theta + v_\mathrm{y}t,\ \cos\theta + v_\mathrm{z}t)$$

is a map $\phi_t^* T : P \to \mathbb{R}$ which assigns a temperature to a point on the planet's surface. In more colorful language we can say that we pulled the temperature field in the cloud back to the surface of the planet. In our example we regarded the manifold P as a subset of C using the embedding ϕ_t. This is a rather common situation. It is often used to relate the internal properties of a manifold to those of the surrounding space.

Now that we have defined the concept of a manifold we need to discuss some simple geometric objects on them which will form the building blocks in expressing physical laws in a coordinate free manner. These building blocks will be vectors and tensors and will be discussed in detail in the next Chapter. After that we will return to discuss physics again.

[11] One could imagine more exotic manifolds with more complicated coordinate maps but this would make our physical example a bit far fetched.

Chapter 3

Vectors and tensors

In this mathematical intermezzo we will define some central geometrical concepts, mainly vectors and tensors, which will form the building blocks for the coordinate independent description of the physical laws in the following Chapters.

3.1 Vectors

We will give a definition of a vector on a general manifold. We start our discussion with a simple example

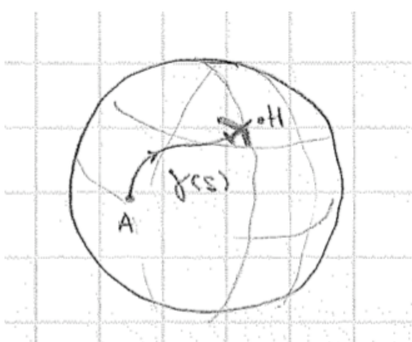

Figure 3.1: Trajectory of a plane on Earth's surface

Imagine a plane flying over the surface of the Earth, say from Amsterdam to Helsinki. The path of the plane is described by a curve $\gamma(s)$, where we let s be the distance to Amsterdam as measured along the curve. Attached to the plane there is thermometer that measures the local outside temperature. Since the temperature is position dependent the temperature field T presents a mapping from points on the Earth's surface to the real numbers $T : M \to \mathbb{R}$, where M is the Earth's surface. The temperature difference between two points on the curve γ (say between Amsterdam and Helsinki) is clearly independent of the coordinate system used to parametrizes the Earth's surface. In particular, we can measure the temperature difference ΔT between two nearby points separated by a distance Δs along the flight path of the plane. Its ratio

$$\frac{\Delta T}{\Delta s} = \frac{T(\gamma(s+\Delta s)) - T(\gamma(s))}{\Delta s}$$

measures a local temperature gradient. If we take the limit $\Delta s \to 0$ we obtain

$$\frac{\partial T}{\partial s}(s) = \lim_{\Delta s \to 0} \frac{T(\gamma(s+\Delta s)) - T(\gamma(s))}{\Delta s}. \tag{3.1}$$

This quantity is clearly coordinate independent. Physically it measures the rate of change of the temperature along the flight path at a distance s measured along the path from Amsterdam. The pilot crew in the plane can measure it without any knowledge of any coordinate system that parametrizes the flight path. Let us now imagine two external observers that want to describe this physical measurement in two different coordinate systems. Since the surface of the Earth is two-dimensional (we assume that the plane flies at constant height) we only need two coordinates, like the pair (θ, ϕ) that represent the latitude and longitude. Let us call the two coordinate systems (x^1, x^2) and (y^1, y^2) (we use superindices for reasons explained later). In the coordinate system where $\gamma(s) = (x^1(s), x^2(s))$ the temperature gradient (7.46) becomes

$$\frac{\partial T}{\partial s}(s) = \frac{\partial T}{\partial x^1}\frac{\partial x^1}{\partial s} + \frac{\partial T}{\partial x^2}\frac{\partial x^2}{\partial s} \tag{3.2}$$

whereas in the other coordinate system we have

$$\frac{\partial T}{\partial s}(s) = \frac{\partial T}{\partial y^1}\frac{\partial y^1}{\partial s} + \frac{\partial T}{\partial y^2}\frac{\partial y^2}{\partial s} \tag{3.3}$$

Since the value of $\partial T/\partial s$ is independent of the coordinate system we can write

$$\frac{\partial T}{\partial x^1}\frac{\partial x^1}{\partial s} + \frac{\partial T}{\partial x^2}\frac{\partial x^2}{\partial s} = \frac{\partial T}{\partial y^1}\frac{\partial y^1}{\partial s} + \frac{\partial T}{\partial y^2}\frac{\partial y^2}{\partial s}.$$

It is clear that this expression is valid no matter what is the form of the temperature field T. We may therefore as well write

$$v(s) = \frac{\partial x^1}{\partial s}\frac{\partial}{\partial x^1} + \frac{\partial x^2}{\partial s}\frac{\partial}{\partial x^2} = \frac{\partial y^1}{\partial s}\frac{\partial}{\partial y^1} + \frac{\partial y^2}{\partial s}\frac{\partial}{\partial y^2} \tag{3.4}$$

where $v(s)$ is a differential operator that can act on any temperature field T. It is, of course, independent of the temperature field and only dependent on the curve $\gamma(s)$. We will call this operator the *tangent vector* at s along the curve $\gamma(s)$. In the (x^1, x^2) coordinate system we can write this vector as

$$v(s) = \frac{\partial x^1}{\partial s}\frac{\partial}{\partial x^1} + \frac{\partial x^2}{\partial s}\frac{\partial}{\partial x^2} = \frac{\partial x^1}{\partial s}e_1 + \frac{\partial x^2}{\partial s}e_2 = \left(\frac{\partial x^1}{\partial s}, \frac{\partial x^2}{\partial s}\right) = (v^1, v^2)$$

where we denoted $e_i = \partial/\partial x^i$ which act as two linearly independent basis vectors. The last two terms between brackets give the standard notation for the vector $v(s)$ as the collection of components with respect to the basis e_i. In particular we have $e_1 = (1, 0)$ and $e_2 = (0, 1)$. In the (y^1, y^2) coordinate system we can write similarly

$$v(s) = \frac{\partial y^1}{\partial s}\frac{\partial}{\partial y^1} + \frac{\partial y^2}{\partial s}\frac{\partial}{\partial y^2} = \left(\frac{\partial y^1}{\partial s}, \frac{\partial y^2}{\partial s}\right) = (w^1, w^2).$$

where the last two brackets give the components with respect to the basis vectors $\partial/\partial y^i$. It is clear from the chain rule of differentiation that the vector components (v^1, v^2) and (w^1, w^2) are simply related by

$$v^i = \frac{\partial x^i}{\partial s} = \sum_{j=1}^{2} \frac{\partial x^i}{\partial y^j}\frac{\partial y^j}{\partial s} = \sum_{j=1}^{2} \frac{\partial x^i}{\partial y^j}\, w^j.$$

Our example motivates the following general definition of a vector on a manifold M. A *vector* v in a point P with coordinate $x = (x^1, \ldots, x^n)$ on a n-dimensional manifold M is a linear operator of the form

$$v = \sum_{i=1}^{n} v^i(x) \frac{\partial}{\partial x^i} \tag{3.5}$$

that acts on functions $T : M \to \mathbb{R}$. If this vector is represented in a different coordinate system $y = (y^1, \ldots, y^n)$, i.e.

$$v = \sum_{i=1}^{n} w^i(y) \frac{\partial}{\partial y^i}$$

then from the chain rule of differentiation

$$\frac{\partial}{\partial x^i} = \sum_{j=1}^{n} \frac{\partial y^j}{\partial x^i} \frac{\partial}{\partial y^j}$$

it follows that

$$\begin{aligned} w^i(y) &= \sum_{j=1}^{n} v^j(x) \frac{\partial y^i}{\partial x^j} && (3.6) \\ v^i(x) &= \sum_{j=1}^{n} w^j(y) \frac{\partial x^i}{\partial y^j} && (3.7) \end{aligned}$$

These two equations were classically used to define vectors, simply as equivalence classes of coefficients transforming according Eqs.(3.6) and (3.7).
It is, however, possible to give an elegant and inherently coordinate independent definition of a vector on a manifold. If $f : M \to \mathbb{R}$ and $g : M \to \mathbb{R}$ are functions on a manifold M, then from Eq.(3.5) it follows that

$$\begin{aligned} v(\alpha f + \beta g)(x) &= \alpha\, v(f)(x) + \beta\, v(g)(x) && (3.8) \\ v(fg)(x) &= f(x)\, v(g)(x) + g(x)\, v(f)(x) && (3.9) \end{aligned}$$

where α and β are real numbers and we introduced the notation

$$v(f)(x) = \sum_{i=1}^{n} v^i(x) \frac{\partial f}{\partial x^i}(x). \tag{3.10}$$

We can now turn the situation around and use Eqs.(3.8) and (3.9) to define vectors on a manifold. Here it is. A *vector* v on a manifold M is a linear operator on functions $f : M \to \mathbb{R}$ having the properties (3.8) and (3.9). If these functions are arbitrarily often differentiable then we can prove that $v(f)$ is necessarily of the form of Eq. (3.10). Since the proof is relatively simple we give it here. Let us start by expanding $f(x)$ in a Taylor series around $x = a$,

$$f(x) = f(a) + \sum_{j=1}^{n} (x^j - a^j) \frac{\partial f}{\partial x^j}(a) + \sum_{j=1}^{n} (x^j - a^j) g_j(x)$$

where the functions g_j have the property that $g_j(a) = 0$. We the apply the linear operator v to both sides of this equation.

$$v(f)(x) = v(f(a)) + \sum_{j=1}^{n} \frac{\partial f}{\partial x^j}(a) v(x^j - a^j) + \sum_{j=1}^{n} v((x^j - a^j) g_j) \tag{3.11}$$

Now from rule (3.9) applied to the constant function equal to 1 it follows that

$$v(1) = v(1 \cdot 1) = 1\, v(1) + 1\, v(1) = 2\, v(1)$$

and hence $v(1) = 0$. Therefore for any constant function α we have $v(\alpha) = \alpha v(1) = 0$. Therefore the first term on the right hand side in Eq.(3.11) vanishes, whereas in the second term we have

$$v(x^j - a^j) = v(x^j) - v(a^j) = v(x^j).$$

If we define $v^j(x) = v(x^j)$ then Eq.(3.11) can be rewritten as

$$v(f)(x) = \sum_{j=1}^{n} v^j(x) \frac{\partial f}{\partial x^j}(a) + \sum_{j=1}^{n} v((x^j - a^j) g_j) \tag{3.12}$$

For the last term in this equation we according to (3.11)

$$v((x^j - a^j) g_j) = (x^j - a^j)\, v(g_j) + g_j(x)\, v(x^j - a^j).$$

If we evaluate this in $x = a$ we see that the right hand side vansihes. By evaluating Eq.(3.12) in $x = a$ we thus obtain

$$v(f)(a) = \sum_{j=1}^{n} v^j(a) \frac{\partial f}{\partial x^j}(a) \tag{3.13}$$

which is exactly what we wanted to prove. It is clear from Eq.(3.13) that the set of vectors in a point p (with coordinate a) forms a n-dimensional vector space. We can therefore write

$$v(a) = \sum_{j=1}^{n} v^j(a) e_j$$

where $e_j = \partial/\partial x^j$ forms a basis of this vector space. We will often denote

$$e_j = \frac{\partial}{\partial x^j} = \partial_j$$

and write Eq.(3.13) as

$$v(a) = \sum_{j=1}^{n} v^j(a) e_j = \sum_{j=1}^{n} v^j(a) \partial_j$$

The vector space of tangent vectors in a point p of a manifold M is called the tangent space in p and is usually denoted by $T_p M$. Let us give an example. We simply consider the transformation from Cartesian coordinates (x^1, x^2) to polar coordinates (r, ϕ) in the two-dimensional plane given by

$$\begin{aligned} x^1 &= r \cos\phi \\ x^2 &= r \sin\phi \end{aligned}$$

The basis vectors (e_r, e_ϕ) in the polar coordinate system are given in terms of the basis vectors (e_1, e_2) of the Cartesian system by

$$\begin{aligned} e_r &= \frac{\partial}{\partial r} = \frac{\partial x^1}{\partial r}\frac{\partial}{\partial x^1} + \frac{\partial x^2}{\partial r}\frac{\partial}{\partial x^2} = \cos\phi \frac{\partial}{\partial x^1} + \sin\phi \frac{\partial}{\partial x^2} = \cos\phi\, e_1 + \sin\phi\, e_2 \\ e_\phi &= \frac{\partial}{\partial \phi} = \frac{\partial x^1}{\partial \phi}\frac{\partial}{\partial x^1} + \frac{\partial x^2}{\partial \phi}\frac{\partial}{\partial x^2} = -r\sin\phi \frac{\partial}{\partial x^1} + r\cos\phi \frac{\partial}{\partial x^2} = -r\sin\phi\, e_1 + r\cos\phi\, e_2 \end{aligned}$$

These vectors are drawn in Fig.3.2. Note that the basis vector e_ϕ has Euclidean length r. In general the basis vectors in the new coordinate system need not be normalized nor be orthogonal.

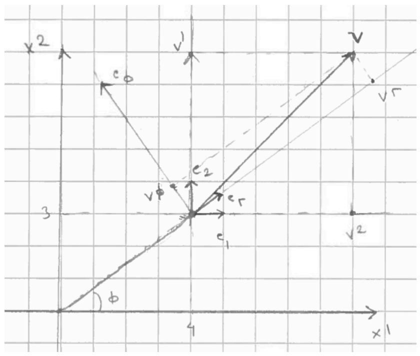

Figure 3.2: A vector v expressed in Cartesian and polar coordinates. In our example the vector is calculated in point $(x^1, x^2) = (4, 3)$ where $v = 5e_1 + 5e_2 = 7e_r + (1/5)e_\phi$.

An arbitrary vector v in the polar coordinate system can therefore be written as

$$v = v^r e_r + v^\phi e_\phi = (v^r \cos\phi - v^\phi r \sin\phi)e_1 + (v^r \sin\phi + v^\phi r \cos\phi)e_2 = v^1 e_1 + v^2 e_2$$

In matrix notation we have

$$\begin{pmatrix} v^1 \\ v^2 \end{pmatrix} = \begin{pmatrix} \cos\phi & -r\sin\phi \\ \sin\phi & r\cos\phi \end{pmatrix} \begin{pmatrix} v^r \\ v^\phi \end{pmatrix} \tag{3.14}$$

or reciprocally

$$\begin{pmatrix} v^r \\ v^\phi \end{pmatrix} = \frac{1}{r} \begin{pmatrix} r\cos\phi & r\sin\phi \\ -\sin\phi & \cos\phi \end{pmatrix} \begin{pmatrix} v^1 \\ v^2 \end{pmatrix} \tag{3.15}$$

these relations are the equivalent of Eqs.(3.6) and (3.7) for our simple example. A graphical picture of a given vector in these two coordinate systems is displayed in Fig.3.2.
The thing to remember from all this analysis is that a vector is not just an arrow with components but a geometrical invariant object. This can, however, be rather confusing at first glance if one is used to draw velocity and force vectors as one does often in physics problems. Let us take the example of a velocity vector, usually written in components as $\mathbf{v} = (v_\mathrm{x}, v_\mathrm{y}, v_\mathrm{z})$ for a particle in some reference frame. We know that if we move along with the particle then in this frame the velocity vector is zero $\mathbf{v}' = (0, 0, 0)$. At first sight this seems in contradiction to Eqs.(3.6) and (3.7) which say that if a vector is zero in one frame, then it is also zero in any other coordinate frame. What we, however, have forgotten is that the transformation between the moving frames also depends on the time coordinate. We have four coordinates (t, x, y, z) and therefore the invariant vector needs four rather than three components. What turns out to be the real invariant geometric concept in the description of the motion of particles is the tangent vector to the world line.
To give an example, let us go back to the example of Fig.3.1. In that example we implicitly assumed in Eq. (3.2) that the temperature field was only dependent on the spatial coordinates (x^1, x^2). However, in general there is also a time-dependence in the temperature field as the temperature will, for instance, change from day to night. To describe this case we should have used the coordinates (t, x^1, x^2) and the flight path is given in these coordinates by

$$\gamma(s) = (t(s), x^1(s), x^2(s)) \tag{3.16}$$

where $t(s)$ is the time value when the plane has travelled a distance s along the path. The path $\gamma(s)$ is therefore simply a world line in space-time. Another parameter, instead of the distance s, to parametrize the path could be the proper time τ passed on a clock in the plane. This is merely a re-parametrization $\tilde{\gamma}(\tau) = \gamma(s(\tau))$ since there is unique relation $s(\tau)$ between the travelled distance s and the proper time τ. Let us, however, stick to the parametrization by s. Then for a time-dependent temperature field the equivalent of Eq.(3.2) becomes

$$\frac{\partial T}{\partial s}(s) = \frac{\partial T}{\partial t}\frac{\partial t}{\partial s} + \frac{\partial T}{\partial x^1}\frac{\partial x^1}{\partial s} + \frac{\partial T}{\partial x^2}\frac{\partial x^2}{\partial s} \tag{3.17}$$

and the tangent vector to the path $\gamma(s)$ becomes

$$v(s) = \frac{\partial t}{\partial s}\frac{\partial}{\partial t} + \frac{\partial x^1}{\partial s}\frac{\partial}{\partial x^1} + \frac{\partial x^2}{\partial s}\frac{\partial}{\partial x^2} = \left(\frac{\partial t}{\partial s}, \frac{\partial x^1}{\partial s}, \frac{\partial x^2}{\partial s}\right) = (v^0, v^1, v^2) \tag{3.18}$$

which is now a tangent vector to world line rather than the tangent vector to a spatial curve. Had we taken the case that the path was parametrized by proper time τ instead of s then, of course, we would have derived Eq.(3.18) with s replaced by τ. In Newtonian mechanics we can always take $\tau = t$ due to the presence of an absolute time and parametrize the flight path as $(t, x^1(t), x^2(t))$ [1]. Let us do this for the motion of a particle in a Cartesian coordinate system with coordinates (t, x, y, z). Then the tangent to the world line is given by

$$v(t) = \frac{d\gamma}{dt} = (1, \frac{dx}{dt}, \frac{dy}{dt}, \frac{dz}{dt}) = (1, \mathbf{v}(t))$$

Under the Galilean transformation $(t', x', y', z') = (t, x - u_{\mathrm{x}}t, y - u_{\mathrm{y}}t, z' - u_{\mathrm{z}}t)$ to a frame moving at relative velocity $\mathbf{u} = (u_{\mathrm{x}}, u_{\mathrm{y}}, u_{\mathrm{z}})$ this four-dimensional vector transforms according to Eq.(3.6) to

$$w = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -u_{\mathrm{x}} & 1 & 0 & 0 \\ -u_{\mathrm{y}} & 0 & 1 & 0 \\ -u_z & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ v_{\mathrm{x}}(t) \\ v_{\mathrm{y}}(t) \\ v_{\mathrm{z}}(t) \end{pmatrix} = \begin{pmatrix} 1 \\ v_{\mathrm{x}}(t) - u_x \\ v_{\mathrm{y}}(t) - u_{\mathrm{y}} \\ v_{\mathrm{z}}(t) - u_{\mathrm{z}} \end{pmatrix}$$

We therefore find the proper addition law for velocities from the vector transformation properties of four-dimensional vectors. In particular, if $\mathbf{v}(t) = 0$ then $v = (1, 0, 0, 0)$ and $w = (1, -\mathbf{u})$. The spatial components of these vectors do not transform according to a vector transformation law, unless we restrict ourselves to purely spatial coordinate transformations.
Let us now investigate how vectors transform under mappings between manifolds. Let $\phi : M \to N$ be a mapping between manifold M and N then we have seen in Eq.(2.63) that ϕ^* pulls back a function f on N to a function $\phi^* f$ on M. If we now have a tangent vector v in p on M then we can assign a vector $\phi_* v$ on N by the definition

$$(\phi_* v)(f) \equiv v(\phi^* f) \tag{3.19}$$

We thus have a mapping $\phi_* : T_p M \to T_{\phi(p)} N$ that maps a vector on M to a vector on N. Since the mapping $\phi : M \to N$ goes in the same direction, we say that ϕ_* describes a *pushforward* of v. Note that we use a subscript asterisk for pushforwards and a superscript for pullbacks.

[1] This is, of course, also possible in a given Lorentz frame in special relativity but there t has no invariant meaning.

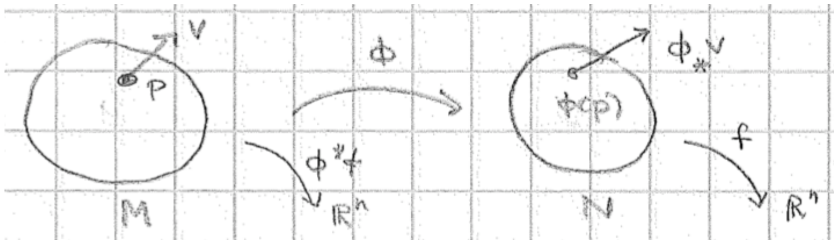

Figure 3.3: Pushing vectors forward to another manifold

The procedure described by Eq.(3.19) looks a bit abstract, but the idea is actually quite simple. Let us illustrate this again with an example. Consider again the flight path of the plane on the Earth's surface. We use the standard longitude and latitude coordinates and the time. More precisely we use the coordinate map

$$\varphi_1(t, R\cos\phi\sin\theta, R\sin\phi\sin\theta, R\cos\theta) = (t, \phi, \theta)$$

to describe an event. In a more mathematical language we would say that every event on the surface of a sphere is part of the manifold $M = \mathbb{R} \times S^2$ (where S^2 is just a common mathematical notation for the two-dimensional surface of a sphere and $\mathbb{R}$ contains the time variables) and that we use the coordinate map $\varphi_1 : M \to \mathbb{R}^3$ to describe these events. The world line of a plane in these coordinates is then given by

$$\gamma(s) = (t(s), \phi(s), \theta(s))$$

where the world line is parametrized by a parameter s with a physical meaning that we can choose (such as a distance or proper time). The tangent vector to this world line is given by

$$v(s) = \frac{d\gamma}{ds} = (\frac{\partial t}{\partial s}, \frac{\partial \phi}{\partial s}, \frac{\partial \theta}{\partial s}) = (v^t(s), v^\phi(s), v^\theta(s))$$

Let us now describe the motion of the plane from a position in space outside the Earth. The outside observer is at rest with respect to the center of mass of the Earth but sees the Earth rotating around its axis. To describe positions in time and space this observer uses four coordinates. We take the corresponding space-time manifold N of the observer simply to be equal to $\mathbb{R}^4$ with the identity map $\varphi_2 : N \to \mathbb{R}^4$ as a coordinate map, i.e.

$$\varphi_2(x^0, x^1, x^2, x^3) = (x^0, x^1, x^2, x^3)$$

We could have used less boring coordinates on $\mathbb{R}^4$ than the Cartesian ones and make the coordinate map φ_2 less trivial but this would not make our example more clear. The world line of the plane in these coordinates is given by

$$\tilde{\gamma}(s) = (x^0(s), x^1(s), x^2(s), x^3(s))$$

How are the coordinates x^j related to the coordinates (t, ϕ, θ)? This is established by a map $\phi : M \to N$ that assigns a point in space to a point on the Earth's surface. If we take into account the uniform rotation of the Earth with angular velocity ω one readily sees that this map is given by

$$\phi(t, R\cos\phi\sin\theta, R\sin\phi\sin\theta, R\cos\theta) = (x^0(t,\phi,\theta), x^1(t,\phi,\theta), x^2(t,\phi,\theta), x^3(t,\phi,\theta)) \quad (3.20)$$

where we defined

$$
\begin{aligned}
x^0(t,\phi,\theta) &= t\\
x^1(t,\phi,\theta) &= R\cos(\phi+\omega t)\sin\theta\\
x^2(t,\phi,\theta) &= R\sin(\phi+\omega t)\sin\theta\\
x^3(t,\phi,\theta) &= R\cos\theta
\end{aligned}
$$

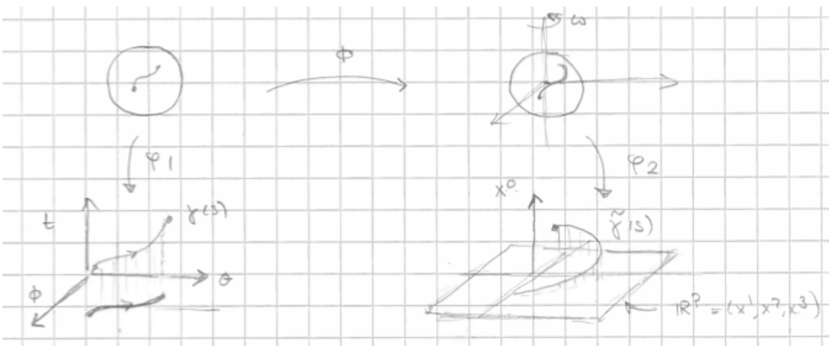

Figure 3.4: World line of a plane in Earth surface coordinates and as seen form an observer in outer space.

With these assignments we have $x^j(s) = x^j(t(s),\phi(s),\theta(s))$ etc. The tangent vector $\tilde{v}$ to the world line $\tilde{\gamma}(s)$ in N is therefore given by

$$\tilde{v}(s) = \frac{\partial\tilde{\gamma}}{\partial s} = (\frac{\partial x^0}{\partial s}, \frac{\partial x^1}{\partial s}, \frac{\partial x^2}{\partial s}, \frac{\partial x^3}{\partial s})$$

Since

$$\frac{\partial x^j}{\partial s} = \frac{\partial x^j}{\partial t}\frac{\partial t}{\partial s} + \frac{\partial x^j}{\partial \phi}\frac{\partial \phi}{\partial s} + \frac{\partial x^j}{\partial \theta}\frac{\partial \theta}{\partial s}$$

we can write

$$\begin{pmatrix} \tilde{v}^0(s) \\ \tilde{v}^1(s) \\ \tilde{v}^2(s) \\ \tilde{v}^3(s) \end{pmatrix} = \begin{pmatrix} \frac{\partial x^0}{\partial t} & \frac{\partial x^0}{\partial \phi} & \frac{\partial x^0}{\partial \theta} \\ \frac{\partial x^1}{\partial t} & \frac{\partial x^1}{\partial \phi} & \frac{\partial x^1}{\partial \theta} \\ \frac{\partial x^2}{\partial t} & \frac{\partial x^2}{\partial \phi} & \frac{\partial x^2}{\partial \theta} \\ \frac{\partial x^3}{\partial t} & \frac{\partial x^3}{\partial \phi} & \frac{\partial x^3}{\partial \theta} \end{pmatrix} \begin{pmatrix} v^t(s) \\ v^\phi(s) \\ v^\theta(s) \end{pmatrix} \tag{3.21}$$

This provides an explicit mapping of a tangent vector v in M to a tangent vector $\tilde{v}$ in N. Let us now see how the same transformation is produced from the definition of a push forward. Let $f : N \to \mathbb{R}$ now be an arbitrary function on N and p an arbitrary point of M with coordinates (t,ϕ,θ) then using the map (3.20) between M and N we have

$$(\phi^* f)(p) = f(\phi(p)) = f(x^0(t,\phi,\theta), x^1(t,\phi,\theta), x^2(t,\phi,\theta), x^3(t,\phi,\theta)) \tag{3.22}$$

Let us now consider an arbitrary vector on M of the form

$$v = v^t\frac{\partial}{\partial t} + v^\phi\frac{\partial}{\partial \phi} + v^\theta\frac{\partial}{\partial \theta} \tag{3.23}$$

Then this vector can be pushed forward to a vector $w = \phi_* v$ on N using Eq.(3.19). We have

$$\begin{aligned} w(f) = (\phi_* v)(f) = v(\phi^* f) &= \left(v^t \frac{\partial}{\partial t} + v^\phi \frac{\partial}{\partial \phi} + v^\theta \frac{\partial}{\partial \theta} \right) f(\phi(p)) \\ &= \sum_{j=0}^{3} \left(v^t \frac{\partial x^j}{\partial t} + v^\phi \frac{\partial x^j}{\partial \phi} + v^\theta \frac{\partial x^j}{\partial \theta} \right) \frac{\partial f}{\partial x^j} = \sum_{j=0}^{3} w^j \frac{\partial f}{\partial x^j} \end{aligned} \tag{3.24}$$

where we defined

$$w^j = v^t \frac{\partial x^j}{\partial t} + v^\phi \frac{\partial x^j}{\partial \phi} + v^\theta \frac{\partial x^j}{\partial \theta}$$

It remains to calculate

$$\begin{aligned} \frac{\partial}{\partial t}(x^0, x^1, x^2, x^3) &= (1, -\omega R \sin(\phi + \omega t) \sin\theta, \omega R \cos(\phi + \omega t) \sin\theta, 0) \\ \frac{\partial}{\partial \phi}(x^0, x^1, x^2, x^3) &= (0, -R \sin(\phi + \omega t) \sin\theta, R \cos(\phi + \omega t) \sin\theta, 0) \\ \frac{\partial}{\partial \theta}(x^0, x^1, x^2, x^3) &= (0, R \cos(\phi + \omega t) \cos\theta, R \sin(\phi + \omega t) \cos\theta, -\sin\theta) \end{aligned}$$

From this we find that we can rewrite Eq.(3.24) in components as

$$\begin{pmatrix} w^0 \\ w^1 \\ w^2 \\ w^3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ -\omega R \sin(\phi + \omega t) \sin\theta & -R \sin(\phi + \omega t) \sin\theta & R \cos(\phi + \omega t) \cos\theta \\ \omega R \cos(\phi + \omega t) \sin\theta & R \cos(\phi + \omega t) \sin\theta & R \sin(\phi + \omega t) \cos\theta \\ 0 & 0 & -\sin\theta \end{pmatrix} \begin{pmatrix} v^t \\ v^\phi \\ v^\theta \end{pmatrix} \tag{3.25}$$

This map tells us exactly how to map an arbitrary vector on M to a vector on N. It is exactly of the same form as Eq.(3.21) but is a generalization of it since it applies to any vector on M not only the tangent vectors to $\gamma(s)$. Let us however, go back to the world line $\gamma(s)$ and choose $s = t$ such that we have the parametrization

$$\gamma(t) = (t, \phi(t), \theta(t))$$

The tangent vector to the world line in M is then given by

$$v(t) = (v^t(t), v^\phi(t), v^\theta(t)) = (1, \frac{\partial \phi}{\partial t}, \frac{\partial \theta}{\partial t})$$

If we insert this expression into the right hand side of Eq.(3.25) and consider the spatial components $\mathbf{w}(t) = (w^1, w^2, w^2)$ of w we recover a three-dimensional vector that describes the velocity of the plane with respect the observer in outer space. In particular, when the plane is at rest on the Earth's surface we have $v(t) = (1, 0, 0)$ and we find that

$$\mathbf{w}(t) = \omega R \sin\theta \begin{pmatrix} -\sin(\phi + \omega t) \\ \cos(\phi + \omega t) \\ 0 \end{pmatrix}$$

and therefore for the space observer the plane moves with constant angular velocity around a circle with radius $R \sin\theta$.

3.2 Metric and volume

3.2.1 Metric

A metric is a generalization of the concept of inner product in a Euclidean space. Inner product itself generalizes the intuitive concept of orthogonality between vectors. Let us imagine a function $g(v,w)$ of vectors v and w which has the property $g(v,w)=0$ whenever two vectors are orthogonal, that is when $v \perp w$. We have not defined what orthogonality is, but it has the following properties

1] $v \perp w \Rightarrow w \perp v$

2] $v \perp w \Rightarrow \alpha v \perp w, \alpha \in \mathbb{R}$

3] $v_1 \perp w$ and $v_2 \perp w \Rightarrow (v_1+v_2) \perp w$

Correspondingly we demand g to satisfy

1] $g(v,w)=g(w,v)$

2] $g(\alpha v,w)=\alpha g(v,w)$

3] $g(v_1+v_2,w)=g(v_1,w)+g(v_2,w)$

Another intuitive property is that if $v \perp w$ for all possible vectors w then v must be the zero vector. This leads to the condition

4] $g(v,w)=0 \quad \forall w \Rightarrow v=0$

This brings us to the following definition. A mapping $g: V \times V \to \mathbb{R}$ that assigns to a pair of vectors (v,w) of a vector space V a real number and satisfies conditions 1]-4] is called a *semi-Riemannian metric*.
If we expand the vectors v and w in a basis then we have

$$g(v,w)=g(\sum_j^n v^j e_j, \sum_k^n w^k e_k)=\sum_{j,k}^n g(e_j,e_k)\, v^j w^k=\sum_{j,k}^n g_{jk}\, v^j w^k$$

where we defined $g_{jk}=g(e_j,e_k)$. The special case $g_{jk}=\delta_{jk}(=1$ if $j=k$ and zero otherwise) gives the standard Euclidean inner product.

$$\langle v,w\rangle=\sum_{jk}^n \delta_{jk}\, v^j w^k = v^1 w^1+\ldots+v^n w^n$$

We then define $v \perp w$ whenever $\langle v,w\rangle=0$. In the case that

$$g_{ij}=\begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

we have

$$g(v,w)=-v^1w^1+v^2w^2+v^3w^3+v^4w^4$$

This is the Minkowski metric that is left invariant by Lorentz transformations. This is essentially the only metric we will use in the remainder of these Lectures. After the discussion of metrics we turn to the next geometric concept, namely the measurement of volumes.

3.2.2 Volume

One concept that is invariant under coordinate transformations is the concept of volume. More precisely we want to assign a number to the volume spanned by n vectors. For instance

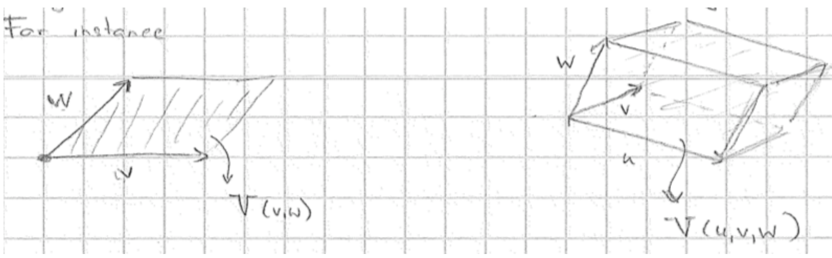

Figure 3.5: The volume spanned by different vectors

In particular, if W is a vector space then we want a map $V(v_1, \dots, v_n)$

$$V : \underbrace{W \times \ldots \times W}_{n \text{ times}} :\to \mathbb{R}$$

that assigns to n vectors the volume spanned by them. Let us list a few of the intuitively desired properties.

1]

$$V(\alpha v_1, v_2, \dots, v_n) = \alpha V(v_1, v_2, \dots, v_n) \quad \alpha \in \mathbb{R} \tag{3.26}$$

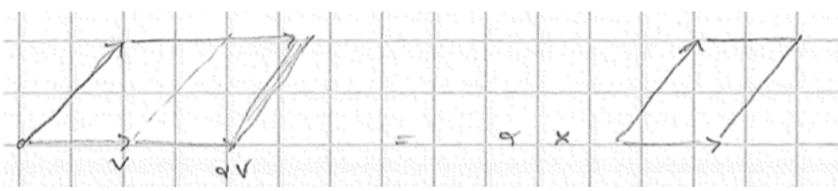

Figure 3.6: Scaling a volume by scaling a spanning vector

where we require this property for the vectors $v_j, j = 2, \dots, n$ as well. If α is negative then the sign of the volume changes. We are not concerned about this since it gives us extra information on the orientation of the vectors. We can always defined $|V|$ to be the volume later.

2]

$$V(u_1 + w_1, v_2, \dots, v_n) = V(u_1, v_2, \dots, v_n) + V(w_1, v_2, \dots, v_n) \tag{3.27}$$

and similarly for $v_2 = u_2 + w_2$ etc. Pictorially

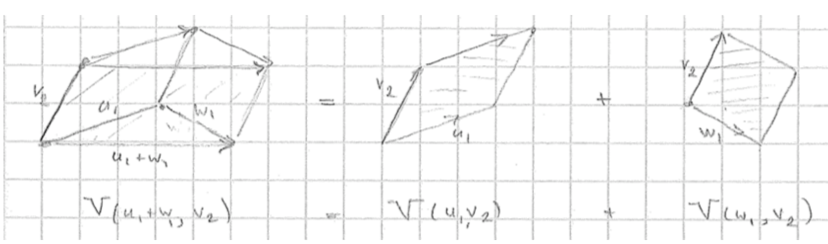

Figure 3.7: Adding to volumes by adding two spanning vectors

Finally, if two vectors are identical then the spanned volume should be zero

3]
$$V(\ldots, v_i, \ldots, v_j, \ldots) = 0 \quad \text{if} \quad v_i = v_j \tag{3.28}$$

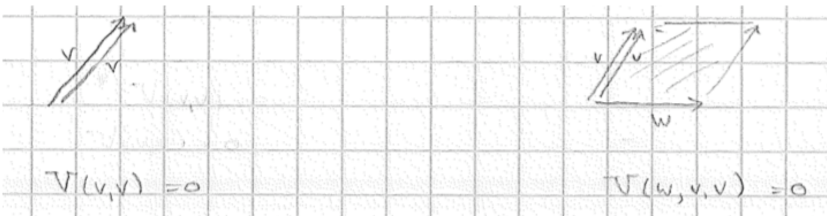

Figure 3.8: The volume spanned by identical vectors is zero

The conditions 1]-3] of Eqs.(3.26)-(3.28) uniquely specify the general form of V. From the conditions (3.27) and (3.28) it follows that

$$\begin{aligned} 0 &= V(u+w, u+w, v_2, \ldots, v_n) = \underbrace{V(u, u, v_2, \ldots, v_n)}_{=0} \\ &+ V(u, w, v_2, \ldots, v_n) + V(w, u, v_2, \ldots, v_n) + \underbrace{V(w, w, v_2, \ldots, v_n)}_{=0} \end{aligned}$$

and hence
$$V(u, w, v_2, \ldots, v_n) = -V(w, u, v_2, \ldots, v_n).$$

We can carry out the same derivation for any two other argument vectors of V and therefore

$$V(\ldots, v_i, \ldots, v_j, \ldots) = -V(\ldots, v_j, \ldots, v_i, \ldots) \tag{3.29}$$

This equation completely fixes the structure of V. If we expand every vector v_i in a basis

$$v_i = \sum_j v_i^j e_j$$

then, according to (3.26), we can write

$$V(v_1, \ldots, v_n) = \sum_{j_1, \ldots, j_n}^{n} v_1^{j_1} v_2^{j_2} \ldots v_n^{j_n} \, V(e_{j_1}, \ldots, e_{j_n}) \tag{3.30}$$

Let us now introduce the convention that the volume spanned by the basis vectors

$$e_1 = (1, 0, \ldots, 0) \quad e_2 = (0, 1, 0, \ldots, 0) \quad \text{etc.}$$

is equal to one, i.e.
$$V(e_1, e_2, \ldots, e_n) = 1. \tag{3.31}$$

Then any interchange of vectors according to Eq.(3.29) introduces a minus sign. If we consider a general permutation $\sigma(1, \ldots, n) = (\sigma(1), \ldots, \sigma(n))$ of the labels $1, \ldots, n$ then from Eq.(3.31) and (3.29) we have that
$$V(e_{\sigma(1)}, \ldots, e_{\sigma(n)}) = (-1)^{|\sigma|} \tag{3.32}$$

where $|\sigma|$ is the number of interchanges required to build the permutation σ. The number $(-1)^{|\sigma|}$ is also known as the *sign* of the permutation. From Eq.(3.28) it also directly follows that none of the indices in Eq.(3.30) need to occur twice since such terms are zero anyway. So the sum can be taken over all possible combinations of different labels and we can therefore write

$$\begin{aligned} V(v_1,\ldots,v_n) &= \sum_\sigma v_1^{\sigma(1)}\ldots v_n^{\sigma(n)} V(e_{\sigma(1)},\ldots,e_{\sigma(n)}) \\ &= \sum_\sigma (-1)^{|\sigma|}\, v_1^{\sigma(1)}\ldots v_n^{\sigma(n)} = \det(v_1,\ldots,v_n) \end{aligned}$$

where the sum over all permutations yields the well-known definition of the determinant. We can therefore write

$$V(v_1,\ldots,v_n) = \begin{vmatrix} v_1^1 & \ldots & v_n^1 \\ \vdots & & \vdots \\ v_1^n & \ldots & v_n^n \end{vmatrix} \tag{3.33}$$

3.3 Tensors

3.3.1 Definition

Both the metric $g(v,w)$ and the volume $V(v_1,\ldots,v_n)$ are multilinear functions acting on vectors, which means that they are mapping from a n-fold product of vector spaces $V\times\ldots\times V$ to the real numbers (or complex numbers if desired)

$$T : \underbrace{V\times\ldots\times V}_{n \text{ times}} \to \mathbb{R} \tag{3.34}$$

with the property that

$$T(v_1,\ldots,\alpha w_j+\beta u_j,\ldots,v_n) = \alpha\, T(v_1,\ldots,w_j,\ldots,v_n) + \beta\, T(v_1,\ldots,u_j,\ldots,v_n) \tag{3.35}$$

Such mappings are called *tensors*.

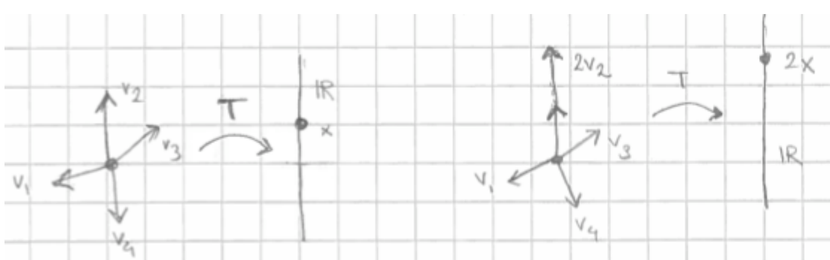

Figure 3.9: A tensor maps a group of vectors to number in a linear way.

If the tensor acts on n vectors we say that it is a tensor of order n. For instance, the metric $g(v_1,v_2)$ is a tensor of order 2 whereas the volume $V(v_1,\ldots,v_n)$ is a tensor of order n. If we expand the vectors v_j in a basis

$$v_j = \sum_{i=1}^m v_j^i e_i \tag{3.36}$$

for a m-dimensional vector space V the from Eq.(3.35) we see that

$$T(v_1,\ldots,v_n) = \sum_{i_1\ldots i_n}^m v_1^{i_1}\ldots v_n^{i_n}\, T(e_{i_1},\ldots,e_{i_n}) = \sum_{i_1\ldots i_n}^m T_{i_1\ldots i_n}\, v_1^{i_1}\ldots v_n^{i_n} \tag{3.37}$$

where we defined the components of the tensor to be

$$T_{i_1 \dots i_n} = T(e_{i_1}, \dots, e_{i_n})$$

Tensors play a crucial role in the theory of electromagnetism. For instance, the Lorentz-invariant relation between force and velocity is described by the electromagnetic tensor F_{ij}, whereas Maxwell's equations consist of two tensorial identities involving exactly the same tensor. Furthermore, energy-momentum conservation laws are described by another tensorial identity. It is therefore important to have a good working knowledge of tensors to follow the remainder of this course.

The most simple tensor is a tensor of order one. According to Eqs. (3.34) and (3.35) this is a linear mapping

$$T : V \to \mathbb{R}$$

with the property

$$T(\alpha v_1 + \beta v_2) = \alpha\, T(v_1) + \beta\, T(v_2).$$

If we take a vector v and expand it in a basis as in Eq.(3.36) we have

$$T(v) = \sum_{j=1}^{m} v^j\, T(e_j) = \sum_{j=1}^{m} T_j\, v^j \tag{3.38}$$

where we defined

$$T_j = T(e_j).$$

Given a basis $\{e_j\}$ in the m-dimensional vector space V we can define m different first order tensors e^j with the property

$$e^j(e_i) = \delta_i^j$$

where $\delta_i^j = 1$ if $i = j$ and zero otherwise. Every first order tensor can now be written as a linear combination of the tensors e^j. This is readily seen. If we write

$$T = \sum_{j=1}^{m} T_j\, e^j \tag{3.39}$$

then

$$T(v) \;=\; \sum_{j=1}^{m} T_j\, e^j \left(\sum_{k=1}^{m} v^k e_k \right) = \sum_{j,k=1}^{m} T_j\, v^k\, e^j(e_k) = \sum_{j,k=1}^{m} T_j\, v^k\, \delta_k^j = \sum_{j=1}^{m} T_j\, v^j$$

which is exactly Eq.(3.38). We thus see from Eq.(3.39) that the tensors e^j form a basis of the linear space of first order tensors on V. This space is called the *dual space* V^* of V. With respect to the basis $\{e^j\}$ of V^* we can therefore write the tensor T in its vector components as

$$T = (T_1, \dots, T_m).$$

The basis $\{e^j\}$ of V^* is called the *dual basis* of the $\{e_j\}$ of V. Our construction immediately raises the question whether we can similarly define a basis in the space of n-th order tensors. This is indeed the case. To do this we start first by defining the tensor product. If T is a tensor of order p and S is a tensor of order q then we define a new tensor $T \otimes S$ of order $p+q$ that acting by

$$(T \otimes S)(v_1, \dots, v_{p+q}) = T(v_1, \dots, v_p) S(v_{p+1}, \dots, v_{p+q}).$$

We have $T \otimes S \neq S \otimes T$ but $S \otimes (T \otimes U) = (S \otimes T) \otimes U$ as is easily checked.

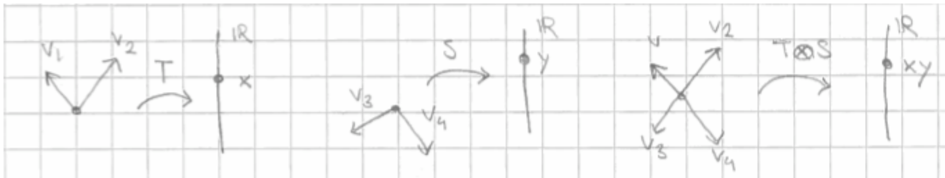

Figure 3.10: A pictorial representation of the tensor product.

With the definition of tensor product the general tensor of Eq.(3.37) can be written as

$$T = \sum_{i_1,\dots,i_n}^{m} T_{i_1\dots i_n}\, e^{i_1}\otimes\dots\otimes e^{i_n} \tag{3.40}$$

Also this expression is readily verified. We have

$$\begin{aligned} T(v_1,\dots,v_n) &= \sum_{i_1,\dots,i_n}^{m} T_{i_1\dots i_n}\, e^{i_1}\otimes\dots\otimes e^{i_n}(v_1,\dots,v_n) \\ &= \sum_{i_1,\dots,i_n}^{m} T_{i_1\dots i_n}\, e^{i_1}(v_1)\dots e^{i_n}(v_n) = \sum_{i_1,\dots,i_n}^{m} T_{i_1\dots i_n}\, v_1^{i_1}\dots v_n^{i_n} \end{aligned}$$

which is exactly Eq.(3.37). We therefore see that the set of n-th order tensors $\{e^{i_1}\otimes\dots\otimes e^{i_n}\}$ forms a basis of the space of all tensors of order n on V. This space will be denoted by $\mathcal{T}^n(V)$. In particular we have $\mathcal{T}^1(V) = V^*$.

Let us now study how the tensor coefficients transform under coordinate transformations. Suppose we had chosen a different basis $\{f_j\}$ in V. This basis is related to the old basis by

$$e_i = \sum_{j}^{m} A_i^j\, f_j. \tag{3.41}$$

Let us then consider again an n-th order tensor T on V. Then in terms of the dual basis $\{f^j\}$ the tensor T has the form

$$T = \sum_{i_1,\dots,i_n}^{m} \tilde{T}_{i_1\dots i_n}\, f^{i_1}\otimes\dots\otimes f^{i_n}. \tag{3.42}$$

From Eq.(3.41) it follows that for any v

$$\begin{aligned} f^i(v) &= \sum_{j}^{m} v^j\, f^i(e_j) = \sum_{j,k}^{m} v^j A_j^k\, f^i(f_k) = \sum_{j,k}^{m} v^j A_j^k\, \delta_k^i \\ &= \sum_{j}^{m} A_j^i\, v^j = \sum_{j}^{m} A_j^i\, e^j(v). \end{aligned}$$

So with this expression we find from Eq.(3.42) that

$$T = \sum_{i_1,\dots,i_n,j_1,\dots j_n}^{m} \tilde{T}_{i_1\dots i_n}\, A_{j_1}^{i_1}\dots A_{j_n}^{i_n}\, e^{j_1}\otimes\dots\otimes e^{j_n}.$$

Comparison to Eq.(3.37) then gives

$$T_{j_1\dots j_n} = \sum_{i_1\dots i_n}^{m} \tilde{T}_{i_1\dots i_n}\, A_{j_1}^{i_1}\dots A_{j_n}^{i_n}.$$

This is how the components of a tensor transform under a coordinate transformation. The tensor is defined independent of a basis as is, for instance, clear for the example of the volume tensor which measures the volume spanned by n independent vectors independent of the coordinate system in which they are expressed.

Since the dual space V^* of first order tensors is a vector space we can define a new type of tensor. This tensor is defined to be the multilinear mapping

$$T : \underbrace{V^* \times \ldots \times V^*}_{n \text{ times}} \to \mathbb{R} \tag{3.43}$$

acting on elements $w_1, \ldots, w_n \in V^*$. Each element w in V^* can be written as

$$w = \sum_j^m w_j \, e^j$$

and therefore

$$T(w_1, \ldots, w_n) = \sum_{i_1 \ldots i_n}^m w_{1,i_i} \ldots w_{n,i_n} \, T(e^{i_1}, \ldots, e^{i_n}) = \sum_{i_1 \ldots i_n}^m T^{i_1 \ldots i_n} \, w_{1,i_i} \ldots w_{n,i_n} \tag{3.44}$$

where we defined

$$T^{i_1 \ldots i_n} = T(e^{i_1}, \ldots, e^{i_n}).$$

To distinguish the tensors of type (3.34) from the tensors of type (3.43) we give them different names. A tensor of type (3.34) is called a *covariant* tensor of order n, whereas a tensor of type (3.43) is called a *contravariant* tensor of order n. The simplest contravariant tensor is a contravariant tensor of order 1, i.e. a linear mapping

$$T : V^* \to \mathbb{R}$$

which maps a covariant tensor of order one (or a dual vector) to a number. To be consistent with our earlier definitions the set of tensors of this form should be denoted by $(V^*)^*$, i.e. the dual space of the dual space, which maybe a somewhat confusing concept. Luckily the space $(V^*)^*$ is the same as our original vector space V since we can define the action of a vector $v \in V$ on a dual vector $w \in V^*$ to be

$$v(w) \equiv w(v).$$

and therefore every vector in V can be regarded as a contravariant tensor of order one. In particular, we have for $e_i \in V$ and $e^j \in V^*$ that

$$e_i(e^j) = e^j(e_i) = \delta_i^j$$

and therefore

$$e_i(w) = \sum_j^m w_j \, e_i(e^j) = w_i.$$

As a consequence of this relation we can rewrite the contravariant tensor of Eq.(3.44) as

$$T = \sum_{i_1 \ldots i_n}^m T^{i_1 \ldots i_n} \, e_{i_1} \otimes \ldots \otimes e_{i_n}. \tag{3.45}$$

Let us now see how this tensor transforms under basis transformations. from Eq.(3.41) it follows that

$$f_i = \sum_{j=1}^{m} (A^{-1})_i^j e_j \tag{3.46}$$

where A^{-1} is the inverse matrix defined by

$$\sum_{l=1}^{m} (A^{-1})_l^j A_i^l = \delta_i^j.$$

If we now describe the tensor T in basis $\{f_i\}$

$$T = \sum_{i_1 \ldots i_n}^{m} \tilde{T}^{i_1 \ldots i_n} f_{i_1} \otimes \ldots \otimes f_{i_n}$$

then it follows from Eq.(3.46) that

$$T = \sum_{i_1 \ldots i_n, j_1 \ldots j_n} \tilde{T}^{i_1 \ldots i_n} \, (A^{-1})_{i_1}^{j_1} \ldots (A^{-1})_{i_n}^{j_n} \, e_{j_1} \otimes \ldots \otimes e_{j_n}$$

and we see by comparing to (3.45) that

$$T^{j_1 \ldots j_n} = \sum_{i_1 \ldots i_n}^{m} \tilde{T}^{i_1 \ldots i_n} \, (A^{-1})_{i_1}^{j_1} \ldots (A^{-1})_{i_n}^{j_n}.$$

We thus see that the components of a contravariant tensor transform in an opposite way as compared to the basis transformation (3.41). This is the origin of the words covariant and contravariant.

After having defined the covariant and contravariant tensors we can go to the final generalization by introducing the *mixed* tensor. A tensor that is covariant of order k and contravariant of order l is a multilinear mapping

$$T : \underbrace{V \times \ldots \times V}_{k \text{ times}} \times \underbrace{V^* \times \ldots \times V^*}_{l \text{ times}} \to \mathbb{R}$$

that assigns to k vectors $v_1, \ldots, v_k \in V$ and l dual vectors $w_1, \ldots, w_l \in V^*$ the real number $T(v_1, \ldots, v_k, w_1, \ldots, w_l)$. In terms of a basis $\{e_i\}$ in V and its dual basis $\{e^i\}$ in V^* this tensor can be written as

$$T = \sum_{i_1 \ldots i_k, j_1 \ldots j_l}^{m} T_{i_1 \ldots i_k}^{j_1 \ldots j_l} \, e^{i_1} \otimes \ldots \otimes e^{i_k} \otimes e_{j_1} \otimes \ldots \otimes e_{j_l} \tag{3.47}$$

The transformation rule under a basis transformation for a mixed tensor is derived completely analogously as that for the co- and contravariant tensors. If in a new basis $\{f_i\}$ in V and its corresponding dual basis we have

$$T = \sum_{i_1 \ldots i_k, j_1 \ldots j_l}^{m} \tilde{T}_{i_1 \ldots i_k}^{j_1 \ldots j_l} \, f^{i_1} \otimes \ldots \otimes f^{i_k} \otimes f_{j_1} \otimes \ldots \otimes f_{j_l}$$

then we have that

$$T_{i_1 \ldots i_k}^{j_1 \ldots j_l} = \sum_{\beta_1 \ldots \beta_k, \alpha_1 \ldots \alpha_l}^{m} T_{\beta_1 \ldots \beta_k}^{\alpha_1 \ldots \alpha_l} \, A_{i_1}^{\beta_1} \ldots A_{i_k}^{\beta_k} (A^{-1})_{\alpha_1}^{j_1} \ldots (A^{-1})_{\alpha_l}^{j_l}$$

This was how historically a tensor was first defined, as objects with indices that transform under basis transformations in this way. The space of tensors on a vector space V that are covariant of order k and contravariant of order l is denoted by $\mathcal{T}_l^k(V)$.

3.3.2 Operations on tensors

Now that we have defined general tensors we can define some commonly used operations on them. One of the simplest ones is the *contraction* . A contraction with respect to indices p and q is a mapping

$$c^p_q : \mathcal{T}^k_l(V) \to \mathcal{T}^{k-1}_{l-1}(V)$$

which is most easily defined in terms of the tensor components. We define

$$\tilde{T}^{j_1 \dots j_{p-1} j_{p+1} \dots j_l}_{i_1 \dots i_{q-1} i_{q+1} \dots i_k} \equiv \sum_{r=1}^{m} T^{j_1 \dots j_{p-1}\, r\, j_{p+1} \dots j_l}_{i_1 \dots i_{q-1}\, r\, i_{q+1} \dots i_k} \tag{3.48}$$

For instance, the contraction $c^1_1 : \mathcal{T}^3_2(V) \to \mathcal{T}^2_1(V)$ is defined by

$$\tilde{T}^{j_2 j_3}_{i_2} = \sum_{r=1}^{m} T^{r\, j_2 j_3}_{r\, i_2}$$

We defined this operation in a basis but as an exercise you can easily convince yourself that the definition (3.48) is indeed basis independent.
The next operation that we define on tensors is usually referred to as the "raising and lowering of indices". We have established a mapping $V \to V^*$ from a vector space to its dual by $e_i \to e^{*i}$ after choosing a basis. However, this mapping depends on the choice of a basis. A basis-independent mapping can be defined when our vector space V is equipped with a metric $g(u, v)$. Being a second rank covariant tensor the metric tensor g can be written as

$$g = \sum_{ij}^{m} g_{ij}\, e^i \otimes e^j$$

where $g_{ij} = g(e_i, e_j)$. We further define a corresponding contra variant tensor g^* by

$$g^* = \sum_{ij}^{m} g^{ij}\, e_i \otimes e_j \tag{3.49}$$

where the matrix g^{ij} is the inverse of g_{ij}, i.e.

$$\sum_{j=1}^{m} g^{ij} g_{jk} = \delta^j_k.$$

(we show below that the inverse exists). Using the metric tensor we can define a mapping from V to V^* as follows. To every $v \in V$ we assign a dual vector $v^\flat \in V^*$ acting on vectors $u \in V$ as follows

$$v^\flat(u) = g(v, u). \tag{3.50}$$

This defines a mapping $\flat : V \to V^*$ which assigns $v^\flat$ to v which is clearly independent of the choice of a basis. Since the dual vector $v^\flat$ (pronounced "v flat") is in V^* it can be expanded in a dual basis as

$$v^\flat = \sum_{j=1}^{n} v^\flat_j\, e^j$$

where $v^\flat_j = v^\flat(e_j)$. This coefficient is readily calculated from Eq.(3.50). We have

$$v^\flat_j = v^\flat(e_j) = g(v, e_j) = \sum_i v^i\, g(e_i, e_j) = \sum_i v^i\, g_{ij}$$

Since the location of the upper or lower index indicates that we deal with a vector or its dual one often simply writes $v^\flat_j = v_j$ such that

$$v^\flat_j = v_j = \sum_i v^i\, g_{ij}$$

One says that the dual vector is obtained by lowering the index with the metric tensor. A natural question to ask is whether this mapping is invertible. This is readily shown to be a consequence of the non degeneracy of the metric. Suppose that for two vectors v_1 and v_2 in V we have $v^\flat_1 = v^\flat_2$. This would imply that for all vectors u in V

$$g(v_1, u) = v^\flat_1(u) = v^\flat_2(u) = g(v_2, u)$$

and therefore

$$g(v_1 - v_2, u) = 0.$$

The non-degeneracy of the metric then tells us that $v_1 - v_2 = 0$ or $v_1 = v_2$. This means that the mapping is indeed invertible. The next question we can ask is whether for any $w \in V^*$ there is a vector $v \in V$ such that $w = v^\flat$. This is indeed the case and follows also from the non-degeneracy of the metric. Let us start by taking m linearly independent vectors $v_j \in V$ for $j = 1, \ldots, m$. and construct the m duals $v^\flat_j$. We first show that the dual vectors $v^\flat_j$ are a basis in V^* and hence and $w \in V^*$ can be written as a linear combination of them. Let α^j be a set of m coefficients and suppose that

$$0 = \sum_{j=1}^m \alpha^j\, v^\flat_j \tag{3.51}$$

As a consequence of the definition of $v^\flat_j$ this implies that

$$0 = \sum_{j=1}^m \alpha^j\, v^\flat_j(u) = \sum_{j=1}^m \alpha^j\, g(v_j, u) = g(\sum_{j=1}^m \alpha^j\, v_j, u) \tag{3.52}$$

Since g is non-degenerate this is only possible for all u when

$$0 = \sum_{j=1}^m \alpha^j\, v_j.$$

However, since the vectors v_j where chosen to be linearly independent this is only possible when $\alpha^j = 0$. It therefore follows that Eq.(3.51) can only be valid when all coefficients vanish. But this implies that the dual vectors $v^\flat_j$ form a basis for V^*. Therefore any $w \in V^*$ can be written as a linear combination of them, i.e.

$$w(u) = \sum_{j=1}^m \beta^j\, v^\flat_j(u) = \sum_{j=1}^m \beta^j\, g(v_j, u) = g(\sum_{j=1}^m \beta^j v_j, u)$$

We therefore see that for any $w \in V^*$ there is a unique vector $v \in V$ such that $w = v^\flat$. Let us denote this vector associated to w by $w^\sharp$ (pronounced "w sharp"). More precisely, it is defined by

$$w(u) = g(w^\sharp, u).$$

This defines a mapping $\sharp : V^* \to V$ which assigns $w^\sharp$ to w. If we expand $w^\sharp$ in a basis in V^* we have

$$w^\sharp = \sum_{j=1}^m w^{\sharp,j} e_j,$$

and therefore

$$w_i = w(e_i) = g(w^\sharp, e_i) = \sum_{j=1}^m w^{\sharp,j} g(e_j, e_i) = \sum_{j=1}^m w^{\sharp,j} g_{ji}$$

Multiplying both sides with the inverse matrix g^{ik} (which we now know exists since we proved that $w^\sharp$ is well-defined) of Eq.(3.49) and summing over i then gives

$$w^{\sharp,k} = \sum_{i=1}^m w_i \, g^{ik} \tag{3.53}$$

Again since the location of the indices tells whether we deal with vectors or their duals one often writes $w^{\sharp,k} = w^j$. The vector $w^\sharp$ with coefficient w^j is then obtained by raising the coefficients of the dual vector with the inverse metric. The mappings $\sharp$ and $\flat$ between V and V^* are often called the "musical isomorfisms". Rather than raising and lowering tones as in music they raise and lower indices of tensors. We will now extend the raising and lowering operations to general tensors. We will start with a few examples. Let us consider a rank two covariant tensor $T \in \mathcal{T}_0^2(V)$ which has the form

$$T = \sum_{ij}^m T_{ij} \, e^i \otimes e^j$$

To this tensor we can assign a mixed tensor $\tilde{T} \in \mathcal{T}_1^1(V)$ acting on a vector u and a dual vector w by the definition

$$\tilde{T}(u, w) = T(u, w^\sharp) \tag{3.54}$$

The mixed tensor is of the form

$$\tilde{T} = \sum_{ij}^m \tilde{T}_i^j \, e^i \otimes e_j$$

We therefore have in a basis

$$\sum_{ij} \tilde{T}_i^j u^i w_j = \tilde{T}(u, w) = T(u, w^\sharp) = \sum_{ik} T_{ik} u^i w^{\sharp,k} = \sum_{ijk} T_{ik} u^i g^{kj} w_j$$

and we therefore see that

$$\tilde{T}_i^j = \sum_k T_{ik} g^{kj}$$

We therefore see that we have raised the second index of T by the inverse metric tensor. This was to be expected since the symbol sharp $\sharp$ appeared in the second argument on the right hand side of Eq.(3.54). In a second example we will lower an index. Consider a third order mixed tensor $T \in \mathcal{T}_2^1(V)$ of the form

$$T = \sum_{ijk} T_i^{jk} \, e^i \otimes e_j \otimes e_k$$

We then assign a new tensor $\tilde{T} \in \mathcal{T}_1^2(V)$ by

$$\tilde{T}(u, v, w) = T(u, v^\flat, w)$$

where $u, v \in V$ and $v^\flat, w \in V^*$. The transformation of the tensor coefficients follows immediately by writing out this equation in a basis

$$\sum_{ijk} \tilde{T}_{ij}^k u^i v^j w_k = \tilde{T}(u, v, w) = T(u, v^\flat, w) = \sum_{ilk} T_i^{lk} u^i v_l^\flat w_k = \sum_{ijlk} T_i^{lk} u^i g_{lj} v^j w_k$$

and we therefore find that

$$\tilde{T}^k_{ij} = \sum_l T^{lk}_i g_{lj} \tag{3.55}$$

We therefore lowered an upper index to a lower index using the metric tensor. We note that when raising and lowering indices it is important to keep track of the order of the indices (e.g. is T^p_q obtained from T_{pq} or T_{qp} after raising an index?). We therefore invent a suitable notation. We illustrate it with an example. Let $T \in \mathcal{T}^2_3(V)$ be a mixed tensor acting on two vectors $v_1, v_2 \in V$ and three dual vectors $w_1, w_2, w_3 \in V^*$, but not in the order of Eq.(3.47). Instead we have the order

$$T(v_1, w_1, w_2, v_2, w_3) \tag{3.56}$$

If we write this out in a basis we write

$$T = \sum_{i_1 \ldots i_5}^m T_{i_1}{}^{i_2 i_3}{}_{i_4}{}^{i_5}\, e^{i_1} \otimes e_{i_2} \otimes e_{i_3} \otimes e^{i_4} \otimes e_{i_5}$$

i.e. the order of the indices up and down correspond to the order in which the arguments in Eq.(3.56) appear. With this notation the tensor of Eq.(3.47) is written as

$$T = \sum_{i_1 \ldots i_k, j_1 \ldots j_l}^m T_{i_1 \ldots i_k}{}^{j_1 \ldots j_l}\, e^{i_1} \otimes \ldots \otimes e^{i_k} \otimes e_{j_1} \otimes \ldots \otimes e_{j_l}$$

With these preliminaries the raising and lowering operation on a general tensor is then described as follows. If $T \in \mathcal{T}^p_q(V)$ we can define assign a new tensor $\tilde{T} \in \mathcal{T}^{p-1}_{q+1}(V)$ by the following procedure. If T acts on a vector as its j-th argument. Then we define

$$\tilde{T}(w) = T(w^\sharp)$$

for that argument and leave all the other arguments untouched (they are suppressed in the notation). This will raise the j-th index of the tensor T. This corresponds to our first example of Eq.(3.54) . In our new notation we have for the indices in this example

$$\tilde{T}_i{}^j = \sum_{k=1}^m T_{ik} g^{kj}$$

where we raised the second index. If had chosen to define $\tilde{T}(w, u) = T(w^\sharp, u)$ instead we would have obtained

$$\tilde{T}^i{}_j = \sum_{k=1}^m T_{kj} g^{ki}$$

which raises the first index. The procedure for lowering an index is analogous. If $T \in \mathcal{T}^p_q(V)$ we can define assign a new tensor $\tilde{T} \in \mathcal{T}^{p+1}_{q-1}(V)$ by the following procedure. If T acts on a dual vector in its j-th argument we define

$$\tilde{T}(v) = T(v^\flat)$$

for that particular argument and leave all other arguments unchanged. This will lower the j-th index of the tensor T. This is what we did in our second example of Eq.(3.55). Being careful with the order of the indices we can write this in our new notation as

$$\tilde{T}_{ij}{}^k = \sum_l T_i{}^{lk} g_{lj}$$

from which it is clear that we lowered the second index.

Now that we have defined the operation of raising and lowering indices it is clear that we can also apply it repeatedly. For instance,

$$T^{ij} = \sum_{pq}^{m} g^{ip} g^{jq}\, T_{pq} \tag{3.57}$$

which in our basis-independent notation amounts to

$$\tilde{T}(w_1, w_2) = T(w_1^\sharp, w_2^\sharp) \tag{3.58}$$

for two dual vectors w_1 and w_2. In the component notation of Eq.(3.57) we removed the tilde on the left hand side of the components. This is customary notation as the upper indices already indicate that we deal with a different tensor. An interesting special case is obtained when we take $T = g$ equal to the metric tensor. In that case Eq.(3.57) becomes the identity

$$g^{ij} = \sum_{pq}^{m} g^{ip} g^{jq}\, g_{pq} = \sum_{p} g^{ip} \delta_p^j = g^{ij} \tag{3.59}$$

where we used $g_{ij} = g_{ji}$. We further see from Eq.(3.58) that we have the identity

$$g^*(w_1, w_2) = g(w_1^\sharp, w_2^\sharp)$$

for two dual vectors w_1 and w_2.

3.3.3 Properties of the metric

The metric tensor presents a non-degenerate metric. In Appendix A we will show that for any non-degenerate metric on a m-dimensional vector space V we can always find a basis e_j in V such that

$$g_{ij} = g(e_i, e_j) = \begin{cases} \pm 1 & \text{if } i = j \text{ for } i = 1, \ldots, m \\ 0 & \text{otherwise} \end{cases}$$

Such a basis will be called an orthonormal basis for the metric g. For the case of a Minkowski metric on a 4-dimensional vector space the metric looks as follows

$$g_{ij} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

In this case (and in fact for any non-degenerate metric) we can distinguish three types of vectors v:

$g(v, v) > 0$	space-like vectors
$g(v, v) = 0$	light-like vectors
$g(v, v) < 0$	time-like vectors

The vectors with $g(v, v) > 0 (< 0)$ point more in the space (time) direction than in the time (space) direction and are called space- (time-) like vectors. Vectors with the property $g(v, v) = 0$ describe the propagation of light rays and are hence called light-like.

A vector space with metric g can have a completely light-like basis. For instance, for our Minkowski metric in terms of an orthonormal basis e_1, e_2, e_3, e_4 with

$$1 = -g(e_1, e_1) = g(e_2, e_2) = g(e_3, e_3) = g(e_4, e_4)$$

and $g(e_i, e_j) = 0$ for $i \neq j$ we can construct a new basis u_1, u_2, u_3, u_4 given by

$$(u_1, u_2, u_3, u_4) = (e_1 - e_2, e_1 + e_2, e_1 + e_3, e_1 + e_4)$$

One can check that all u_i are light-like

$$g(u_i, u_i) = 0$$

The vectors u_i are linearly independent and form a basis. We can, for instance, express the original basis e_i in terms of the new basis as follows

$$(e_1, e_2, e_3, e_4) = (\frac{1}{2}(u_1 + u_2), \frac{1}{2}(u_2 - u_1), u_3 - \frac{1}{2}u_1 - \frac{1}{2}u_2, u_4 - \frac{1}{2}u_1 - \frac{1}{2}u_2)$$

Hence any other vector in V can be expressed in terms of the vectors u_i. The existence of purely light-like bases means that the standard proof for the existence of an orthonormal basis for positive definite inner products can not carried over directly to non positive definite inner products such as the Minkowski metric. The standard proof is based on the Gram-Schmidt procedure which assumes any vector to be normalizable to one, which clearly fails for light-like vectors. The existence of a basis can nevertheless be proven and the proof is presented in Appendix A.

3.3.4 A metric on the space of tensors

The metric g can be regarded as defining a non positive-definite inner product between vectors. For any two vectors u and v in V we can define

$$(u|v) \equiv g(u, v) = \sum_{ij}^{m} g_{ij} u^i v^j$$

We can define such an inner product for dual vectors $w_1, w_2 \in V^*$ as well by defining

$$(w_1|w_2) = g^*(w_1, w_2) = \sum_{ij}^{m} g^{ij} w_{1,i} w_{2,j} \tag{3.60}$$

After our discussion on the raising and lowering of tensor indices in the previous section it is not difficult to construct a generalization of these two equations which provides an inner product between general tensors. For instance, for tensors $T, S \in \mathcal{T}_0^2(V)$ we can define

$$(T|S) = \sum_{i_1 i_2}^{m} T_{i_1 i_2} \tilde{S}^{i_1 i_2} = \sum_{i_1 i_2 j_1 j_2} T_{i_1 i_2} S_{j_1 j_2} g^{i_1 j_1} g^{i_2 j_2}$$

This brings us to the following general definition. For $T, S \in \mathcal{T}_q^p(V)$ we define

$$(T|S) = \sum_{i_1 \ldots i_p, j_1 \ldots j_q}^{m} T^{j_1 \ldots j_q}_{i_1 \ldots i_p} \tilde{S}^{i_1 \ldots i_p}_{j_1 \ldots j_q} \tag{3.61}$$

where

$$\tilde{S}^{i_1 \ldots i_p}_{j_1 \ldots j_q} = \sum_{k_1 \ldots k_p, l_1 \ldots l_q}^{m} S^{l_1 \ldots l_q}_{k_1 \ldots k_p} g^{i_1 k_1} \ldots g^{i_p k_p} g_{l_1 j_1} \ldots g_{l_q j_q}$$

by raising and lowering of all the indices. It is not difficult to check the following properties of this inner product:

1. $(T|S) = (S|T)$
2. $(\alpha_1 T_1 + \alpha_2 T_2|S) = \alpha_1(T_1|S) + \alpha_2(T_2|S)$ for α_1, α_2 real numbers and a similar linearity applies to the second argument S.
3. If $(T|S) = 0$ for all $S \in \mathcal{T}^p_q(V)$ then $T = 0$

These properties tell us that Eq.(3.61) defines a non-degenerate metric on the space of all tensors. From the previous section we know that any non-degenerate metric allows for the existence of an orthonormal basis. We therefore conclude that there is a basis of tensors $e_i \in \mathcal{T}^p_q(V)$ with the property

$$\begin{aligned} (e_i|e_j) &= \pm 1 \quad \text{for } i = 1, \ldots, m(p+q) \\ (e_i|e_j) &= 0 \quad \text{for } i \neq j \end{aligned}$$

The simplest case of a tensor inner product is, of course, the metric itself for which

$$g_{ij} = g(e_i, e_j) = (e_i|e_j)$$

and hence if e_i is an orthonormal basis for the metric then it is trivially also an orthonormal basis for the tensor inner product. The next simplest case is that of the inner product of dual vectors as in Eq.(3.60). If e^j is a dual basis to e_i then since $w_j = w(e_j)$ we can write Eq.(3.60) as

$$(w_1|w_2) = \sum_{ij}^{m} g^{ij} w_1(e_i) w_2(e_j)$$

and in particular

$$(e^k|e^l) = \sum_{ij}^{m} g^{ij} e^k(e_i) e^l(e_j) = \sum_{ij}^{m} g^{ij} \delta^k_i \delta^l_j = g^{kl}$$

In case that e^k was dual to an orthornormal basis we have that g is diagonal and $g^{ii} = g_{ii} = \pm 1$ and hence

$$(e^k|e^l) = \begin{cases} \pm 1 & \text{if } k = l \text{ for } k = 1, \ldots, m \\ 0 & \text{otherwise} \end{cases}$$

Let us now consider two special p-th order covariant tensors of the form

$$T = v_1 \otimes \ldots \otimes v_p \qquad S = u_1 \otimes \ldots \otimes u_p$$

where $u_i, v_j \in V^*$. Then

$$\begin{aligned} (T|S) &= \sum_{i_1 \ldots i_p, j_1 \ldots j_p}^{m} v_{1,i_1} \ldots v_{p,i_p} g^{i_1 j_1} \ldots g^{i_p j_p} u_{1,j_1} \ldots u_{p,j_p} \\ &= (u_1|v_1) \ldots (u_p|v_p) \end{aligned} \tag{3.62}$$

From this expression it follows that if e_i is an orthonormal basis in V then the tensors $e^{i_1} \otimes \ldots \otimes e^{i_p}$ form an orthonormal basis for the p-th order covariant tensors $\mathcal{T}^p_0(V)$, i.e.

$$\begin{aligned} (e^{i_1} \otimes \ldots \otimes e^{i_p} | e^{j_1} \otimes \ldots \otimes e^{j_p}) &= (e^{i_1}|e^{j_1}) \ldots (e^{i_p}|e^{j_p}) \\ &= \begin{cases} g^{i_1 j_1} \ldots g^{i_p j_p} = \pm 1 & \text{if } (i_1 \ldots i_p) = (j_1 \ldots j_p) \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

For the discussion in the next section it is necessary to pay some extra attention to the anti-symmetric covariant tensors. An arbitrary such tensor can be written as

$$T = \sum_{i_1 \ldots i_p}^{m} T_{i_1 \ldots i_p} \, e^{i_1} \otimes \ldots \otimes e^{i_p}$$

where $T_{i_1 \ldots i_p} = T(e_{i_1}, \ldots, e_{i_p})$. Since T is anti-symmetric we have

$$T(e_{\sigma(1)}, \ldots, e_{\sigma(p)}) = (-1)^{|\sigma|} T(e_1, \ldots, e_p)$$

where σ is a permutation of the indices $(1, \ldots, p)$. For an anti-symmetric tensor S and T we can thus write the inner product as

$$(T|S) = \sum_{i_1 \ldots i_p}^{m} T_{i_1 \ldots i_p} \tilde{S}^{i_1 \ldots i_p} = p! \sum_{i_1 < \ldots < i_p}^{m} T_{i_1 \ldots i_p} \tilde{S}^{i_1 \ldots i_p}$$

where we sum over all ordered p-tuples $i_1 < \ldots < i_p$ of the set $(1, \ldots, m)$. For the anti-symmetric tensors of order p it is convenient to redefine the inner product as

$$\langle T, S \rangle \equiv \frac{1}{p!} (T|S) \tag{3.63}$$

such that

$$\langle T, S \rangle = \sum_{i_1 < \ldots < i_p}^{m} T_{i_1 \ldots i_p} \tilde{S}^{i_1 \ldots i_p}$$

Let us now see what the orthonormal basis is with respect to this inner product. From the anti-symmetry of T it follows that

$$\begin{aligned}
T &= \sum_{i_1 \ldots i_p}^{m} T_{i_1 \ldots i_p} \, e^{i_1} \otimes \ldots \otimes e^{i_p} \\
&= \sum_{i_1 < \ldots < i_p}^{m} T_{i_1 \ldots i_p} \sum_{\sigma} (-1)^{|\sigma|} e^{\sigma(i_1)} \otimes \ldots \otimes e^{\sigma(i_p)} \\
&= \sum_{i_1 < \ldots < i_p}^{m} T_{i_1 \ldots i_p} e^{i_1 \ldots i_p} \qquad (3.64)
\end{aligned}$$

where we defined the basis tensor

$$e^{i_1 \ldots i_p} \equiv \sum_{\sigma} (-1)^{|\sigma|} e^{\sigma(i_1)} \otimes \ldots \otimes e^{\sigma(i_p)}$$

When V is an m-dimensional vector space then there are $\binom{m}{p}$ such basis functions. Let us now calculate the inner product between two such basis tensors. We have, when τ, σ and ρ are permutations and $\tau = \rho \circ \sigma$ that

$$\begin{aligned}
\langle e^{i_1 \ldots i_p}, e^{j_1 \ldots j_p} \rangle &= \frac{1}{p!} \sum_{\sigma, \tau} (-1)^{|\sigma| + |\tau|} (e^{\sigma(i_1)} \otimes \ldots \otimes e^{\sigma(i_p)} | e^{\tau(j_1)} \otimes \ldots \otimes e^{\tau(j_p)}) \\
&= \frac{1}{p!} \sum_{\sigma, \rho} (-1)^{|\sigma|} (-1)^{|\sigma| + |\rho|} (e^{\sigma(i_1)} | e^{\rho \circ \sigma(j_1)}) \ldots (e^{\sigma(i_p)} | e^{\rho \circ \sigma(j_p)}) \\
&= \sum_{\rho} (-1)^{|\rho|} (e^{i_1} | e^{\rho(j_1)}) \ldots (e^{i_p} | e^{\rho(j_p)}) \\
&= \sum_{\rho} (-1)^{|\rho|} g^{i_1 \, \rho(j_1)} \ldots g^{i_p \, \rho(j_p)} = \det(G) \qquad (3.65)
\end{aligned}$$

where G is the matrix with entries $G_{pq} = g^{i_p i_q}$. For instance

$$\langle e^{12}, e^{34} \rangle = \begin{vmatrix} g^{13} & g^{14} \\ g^{23} & g^{24} \end{vmatrix}$$

In case that the basis e_i is orthonormal we have from Eq.(3.65) that

$$\begin{aligned} \langle e^{i_1 \dots i_p}, e^{j_1 \dots j_p} \rangle &= \sum_\rho (-1)^{|\rho|} \, \delta^{i_1}_{\rho(j_1)} \dots \delta^{i_p}_{\rho(j_p)} \, g^{i_1 i_1} \dots g^{i_p i_p} \\ &= \delta^{i_1 \dots i_p}_{j_1 \dots j_p} \, g^{i_1 i_1} \dots g^{i_p i_p} \end{aligned} \tag{3.66}$$

where we defined the symbol

$$\delta^{i_1 \dots i_p}_{j_1 \dots j_p} = \begin{cases} 1 & \text{if } (i_1 \dots i_p) \text{ is an even permutation of } (j_1 \dots j_p) \\ -1 & \text{if } (i_1 \dots i_p) \text{ is an odd permutation of } (j_1 \dots j_p) \\ 0 & \text{otherwise} \end{cases} \tag{3.67}$$

For instance,

$$\delta^{12}_{12} = -\delta^{12}_{21} = 1, \quad \delta^{12}_{13} = 0$$

Since for an orthonormal basis $g^{ii} = \pm 1$ we see that in such a basis the tensors $e^{i_1 \dots i_p}$ for $i_1 < \dots < i_p$ form an $\binom{m}{p}$-dimensional basis for the anti-symmetric covariant tensors of rank p, i.e.

$$\begin{aligned} \langle e^{i_1 \dots i_p}, e^{i_1 \dots i_p} \rangle &= g^{i_1 i_1} \dots g^{i_p i_p} = \pm 1 \\ \langle e^{i_1 \dots i_p}, e^{j_1 \dots j_p} \rangle &= 0 \quad \text{if } (i_1 < \dots < i_p) \neq (j_1 < \dots < j_p) \end{aligned}$$

This concludes our discussion on the inner product of tensors.

3.3.5 The wedge product of tensors

In the theory of electromagnetism and in physics in general anti-symmetric tensors play an important role. For instance, the central quantity in electromagnetism is the anti-symmetric field tensor $F_{\mu\nu}$, which is also closely related to the concept of curvature. Another reason anti-symmetric tensors are important is that they describe volume elements and consequently they play a crucial role in integration of volumes or fluxes. We will describe this in more detail later. For the moment we recall the volume tensor

$$V(v_1, \dots, v_n) = \sum_{j_1 \dots j_n}^{n} v_1^{j_1} \dots v_n^{j_n} V(e_{j_1}, \dots, e_{j_n})$$

where $V(e_{\sigma(1)}, \dots, e_{\sigma(n)}) = (-1)^{|\sigma|}$. If we define

$$\epsilon_{j_1 \dots j_n} = \begin{cases} 1 & \text{if } (j_1 \dots j_n) \text{ is an even permutation of } (1 \dots n) \\ -1 & \text{if } (j_1 \dots j_n) \text{ is an odd permutation of } (1 \dots n) \end{cases}$$

then we can write

$$V(v_1, \dots, v_n) = \sum_{j_1 \dots j_n}^{n} \epsilon_{j_1 \dots j_n} \, v_1^{j_1} \dots v_n^{j_n}$$

or equivalently

$$V = \sum_{j_1 \dots j_n}^{n} \epsilon_{j_1 \dots j_n} \, e^{i_1} \otimes \dots \otimes e^{i_n} \tag{3.68}$$

Let us now look at more general anti-symmetric tensors such as

$$T = \alpha(e^1 \otimes e^2 - e^2 \otimes e^1) + \beta(e^1 \otimes e^3 - e^3 \otimes e^1) + \gamma(e^2 \otimes e^3 - e^3 \otimes e^2) \tag{3.69}$$

It is easy to check that $T(v_1, v_2) = -T(v_2, v_1)$ for all vectors $v_1, v_2 \in V$. If S and T are anti-symmetric tensors then their tensor product is in general not. For instance, if $S \in \mathcal{T}_0^3(V)$ and $T \in \mathcal{T}_0^2(V)$ are both anti-symmetric then

$$(S \otimes T)(v_1, v_2, v_3, v_4, v_5) = S(v_1, v_2, v_3)T(v_4, v_5)$$

is not anti-symmetric upon interchange of the vectors v_1, v_2, v_3 with v_4, v_5. We will therefore introduce a new product $S \wedge T$, known as the *wedge product* that does produce an anti-symmetric tensor. First some definitions.

We define $\Omega^p(V)$ to be the space of p-fold covariant anti-symmetric tensors on a vector space V. We include the case $p = 1$ as well and define $\Omega^1(V) = V^*$. Then a set of indices $(i_1 \dots i_p)$ will be denoted by I and similarly a set of indices $(j_1 \dots j_p)$ will be denoted by J. Further

$$v_I \equiv (v_{i_1}, \dots, v_{i_p})$$

and

$$\delta^I_J \equiv \delta^{i_1 \dots i_p}_{j_1 \dots j_p}$$

where $\delta^{i_1 \dots i_p}_{j_1 \dots j_p}$ was defined before in Eq.(3.67). Then for $S \in \Omega^p(V)$ and $T \in \Omega^q(V)$ we define $S \wedge T \in \Omega^{p+q}(V)$ by

$$(S \wedge T)(v_I) \equiv \frac{1}{p!q!} \sum_{JK} \delta^{JK}_I \, S(v_J) \, T(v_K) \tag{3.70}$$

where $JK = (j_1 \dots j_p, k_1 \dots k_q)$, $J = (j_1 \dots j_p)$ and $K = (k_1 \dots k_q)$ are subsets of I. The sum in Eq. (3.70) runs over all subsets of I. Since δ^{JK}_I changes sign upon interchange of any two indices in I it is clear that the tensor $S \wedge T$ is anti-symmetric and hence element of $\Omega^{p+q}(V)$. Let us give an example. If $S \in \Omega^2(V)$ and $T \in \Omega^1(V) = V^*$ we have

$$\begin{aligned}
(S \wedge T)(v_1, v_2, v_3) = \frac{1}{2!1!}\Big[&\delta^{123}_{123} \, S(v_1, v_2) \, T(v_3) + \delta^{213}_{123} \, S(v_2, v_1) \, T(v_3) \\
&+\delta^{132}_{123} \, S(v_1, v_3) \, T(v_2) + \delta^{312}_{123} \, S(v_3, v_1) \, T(v_2) \\
&+\delta^{231}_{123} \, S(v_2, v_3) \, T(v_1) + \delta^{321}_{123} \, S(v_3, v_2) \, T(v_1)\Big] \\
=&\frac{1}{2}(S(v_1, v_2) - S(v_2, v_1))T(v_3) - \frac{1}{2}(S(v_1, v_3) - S(v_3, v_1))T(v_3) \\
&+ \frac{1}{2}(S(v_2, v_3) - S(v_3, v_2))T(v_1)
\end{aligned} \tag{3.71}$$

This represents an anti-symmetric tensor, even if S would not have been anti-symmetric (and this is also true for the general Eq.(50)) . However, since we know that $S(v_1, v_2) = -S(v_2, v_1)$ we can simplify Eq.(3.71) to

$$(S \wedge T)(v_1, v_2, v_3) = S(v_1, v_2)T(v_3) - S(v_1, v_3)T(v_2) + S(v_2, v_3)T(v_1) \tag{3.72}$$

This procedure works in general. If S and T in Eq.(3.70) are anti-symmetric (and they are by definition) we can rewrite the equation as

$$(S \wedge T)(v_I) = \sum_{\underline{J}\,\underline{K}} \delta^{JK}_I \, S(v_J) \, T(v_K) \tag{3.73}$$

where now $\underline{J} = (j_1 < \ldots < j_p)$ and $\underline{K} = (k_1 < \ldots < k_q)$ are ordered subsets of I. This is easily derived by noting that both δ_I^{JK} and $S(v_J)$ and $T(v_K)$ are anti-symmetric under permutations of indices in J and K. For example, if $S, T \in \Omega^2(V)$ then

$$
\begin{aligned}
(S \wedge T)(v_1, v_2, v_3, v_4) = & \sum_{j_1 < j_2} \sum_{k_1 < k_2} \delta_{1234}^{j_1 j_2 k_1 k_2}\, S(v_{j_1}, v_{j_2})\, T(v_{k_1}, v_{k_2}) \\
= & \delta_{1234}^{1234}\, S(v_1, v_2)\, T(v_3, v_4) + \delta_{1234}^{1324}\, S(v_1, v_3)\, T(v_2, v_4) \\
= & \delta_{1234}^{1423}\, S(v_1, v_4)\, T(v_2, v_3) + \delta_{1234}^{2314}\, S(v_2, v_3)\, T(v_1, v_4) \\
= & \delta_{1234}^{2413}\, S(v_2, v_4)\, T(v_1, v_3) + \delta_{3412}^{1234}\, S(v_3, v_4)\, T(v_1, v_2)
\end{aligned}
$$

$$
\begin{aligned}
= & S(v_1, v_2)\, T(v_3, v_4) - S(v_1, v_3)\, T(v_2, v_4) + S(v_1, v_4)\, T(v_2, v_3) \\
& + S(v_2, v_3)\, T(v_1, v_4) - S(v_2, v_4)\, T(v_1, v_3) + S(v_3, v_4)\, T(v_1, v_2)
\end{aligned} \tag{3.74}
$$

One can again check that this tensor is anti-symmetric and hence an element of $\Omega^4(V)$. We have thus established the mapping

$$\wedge : \Omega^p(V) \times \Omega^q(V) \to \Omega^{p+q}(V)$$

This product is known as the *wedge product*. If $S \in \Omega^p(V)$ and $T \in \Omega^q(V)$ then we see that

$$
\begin{aligned}
(S \wedge T)(v_I) \quad = & \quad \sum_{\underline{J}\,\underline{K}} \delta_I^{JK}\, S(v_J)\, T(v_K) = (-1)^{pq} \sum_{\underline{J}\,\underline{K}} \delta_I^{KJ}\, T(v_K)\, S(v_J) \\
= & \quad (-1)^{pq}\, (T \wedge S)(v_I)
\end{aligned}
$$

since we need pq transpositions to go from JK to KJ. We thus have

$$(S \wedge T) = (-1)^{pq} (T \wedge S) \tag{3.75}$$

We can further check that for $S, T, U \in \Omega^p, \Omega^q, \Omega^r$ that

$$
\begin{aligned}
(S \wedge (T \wedge U))(v_I) = & \sum_{\underline{J}\,\underline{K}} \delta_I^{JK}\, S(v_J)\, (T \wedge U)(v_K) \\
= & \sum_{\underline{J}\,\underline{K},\underline{L},\underline{M}} \delta_I^{JK}\, \delta_K^{LM}\, S(v_J)\, T(v_L)\, U(v_M) = \sum_{\underline{J}\,\underline{L},\underline{M}} \delta_I^{JLM}\, S(v_J)\, T(v_L)\, U(v_M) \\
= & \sum_{\underline{J}\,\underline{L},\underline{M},\underline{N}} \delta_N^{JL}\, \delta_I^{NM}\, S(v_J)\, T(v_L)\, U(v_M) = \sum_{\underline{M},\underline{N}} \delta_I^{NM}\, (S \wedge T)(v_N)\, U(v_M) \\
= & ((S \wedge T) \wedge U)(v_I)
\end{aligned}
$$

where we used that

$$\sum_{\underline{K}} \delta_I^{JK}\, \delta_K^{LM} = \delta_I^{JLM}$$

which is an identity that you can check for yourself. We have thus shown that

$$S \wedge (T \wedge U) = (S \wedge T) \wedge U.$$

There is therefore no need to use brackets and we can simply write $S \wedge T \wedge U$. We can therefore apply the wedge product repeatedly. In particular when $T_j \in \Omega^1(V) = V^*$ for $j = 1 \ldots n$ then we can calculate that

$$
\begin{aligned}
(T_1 \wedge \ldots \wedge T_n)(v_1, \ldots, v_n) = & \sum_{j_1 \ldots j_n} \delta_{1 \ldots n}^{j_1 \ldots j_n}\, T_1(v_{j_1}) \ldots T_n(v_{j_n}) \\
= & \sum_{\sigma} (-1)^{|\sigma|}\, T_1(v_{\sigma(1)}) \ldots T_n(v_{\sigma(n)}) = \sum_{\sigma} (-1)^{|\sigma|}\, T_{\sigma(1)}(v_1) \ldots T_{\sigma(n)}(v_n)
\end{aligned}
$$

where we sum over all permutations σ of the labels and therefore

$$T_1 \wedge \ldots \wedge T_n = \sum_\sigma (-1)^{|\sigma|}\, T_{\sigma(1)} \otimes \ldots \otimes T_{\sigma(n)}.$$

If we take the special case $T_k = e^{i_k}$, i.e. the dual basis vectors, then we have

$$e^{i_1} \wedge \ldots \wedge e^{i_n} = \sum_\sigma (-1)^{|\sigma|}\, e^{\sigma(i_1)} \otimes \ldots \otimes e^{\sigma(i_n)} \tag{3.76}$$

We see that these are exactly the basis vectors $e^{i_1 \cdots i_n}$ encountered in Eq.(3.3.4). We can therefore express a general anti-symmetric tensor as in Eq.(3.64) as

$$T = \sum_{i_1 < \ldots < i_p} T_{i_1 \ldots i_p}\, e^{i_1} \wedge \ldots \wedge e^{i_p} \tag{3.77}$$

We have, for instance, from Eq.(3.76) that

$$e^1 \wedge e^2 = e^1 \otimes e^2 - e^2 \otimes e^1$$

and therefore the antisymmetric tensor of Eq.(3.69) can be rewritten as

$$T = \alpha\, e^1 \wedge e^2 + \beta\, e^1 \wedge e^3 + \gamma\, e^2 \wedge e^3$$

With the short notation $e^I = e^{i_1} \wedge \ldots \wedge e^{i_p}$, $T_I = T_{i_1 \ldots i_p}$ and $\underline{I} = (1_1 < \ldots < i_p)$ we can write Eq.(3.77) compactly as

$$T = \sum_{\underline{I}} T_I\, e^I$$

If the dimension of the vector space V is m then the number of basis functions e^I is given by $\binom{m}{p}$ since to construct all e^I we need to pick p distinct integers out of m. We this have that

$$\dim \Omega^p(V) = \binom{m}{p} = \frac{m!}{p!(m-p)!} \qquad m = \dim V \tag{3.78}$$

Let us now think a bit about the geometrical meaning of the wedge product. If we take $u, w \in V^*$ then we have

$$\begin{aligned} u \wedge w &= \left(\sum_i^m u_i e^i \right) \wedge \left(\sum_j^m w_j e^j \right) = \sum_{ij}^m u_i\, w_j\, e^i \wedge e^j \\ &= \sum_{i<j}^m (u_i w_j - u_j w_i)\, e^i \wedge e^j = \sum_{i<j}^m \begin{vmatrix} u_i & w_i \\ u_j & w_j \end{vmatrix} e^i \wedge e^j \end{aligned} \tag{3.79}$$

Now the 2×2-determinant in this equation describes the area in the (e^i, e^j)-plane spanned by the vectors u and w projected on that plane. Pictorially, when V^* is three-dimensional

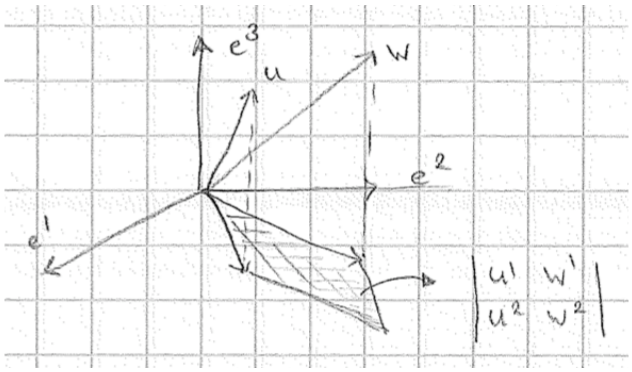

Figure 3.11: The are spanned by vectors u and w in the (e^1, e^2)-plane

One has, however, to be a bit careful with this picture since we have not assumed anything about the orthogonality of e^1, e^2, e^3 which requires a metric to define. We will, however, often take an orthonormal basis. More generally, we have

$$\begin{aligned} w_1 \wedge \ldots \wedge w_p &= \sum_{i_1 \ldots i_p} w_{1,i_1} \ldots w_{p,i_p} \, e^{i_1} \wedge \ldots \wedge e^{i_p} \\ &= \sum_{i_1 < \ldots < i_p} \sum_{\sigma} (-1)^{|\sigma|} \, w_{1,\sigma(i_1)} \ldots w_{n,\sigma(i_p)} \, e^{i_1} \wedge \ldots \wedge e^{i_p} \\ &= \sum_{i_1 < \ldots < i_p} \det(w_{i_1}, \ldots, w_{i_p}) \, e^{i_1} \wedge \ldots \wedge e^{i_p} \end{aligned} \tag{3.80}$$

where $\det(w_{i_1}, \ldots, w_{i_p})$ is the $p \times p$-determinant of vectors $w_{i_1}, \ldots, w_{i_p} \in V^*$ with only rows $(i_1, \ldots, i_p)$ present. This is precisely the p-dimensional volume spanned by the vectors $w_{i_1}, \ldots, w_{i_p}$ projected on the $(e^{i_1}, \ldots, e^{i_p})$-plane of the m-dimensional space V^*. This feature explains why the anti-symmetric tensors play an important role in the theory of integration.

3.3.6 The invariant volume

In Eq.(3.68) we presented the volume tensor. In our new wedge notation this expression can now be written as

$$V = e^1 \wedge \ldots \wedge e^n$$

Let us now consider the transformation to another dual basis f^j of V^*, i.e.

$$e^i = \sum_{j=1}^{n} B^i_j f^j$$

Then the volume tensor V becomes

$$\begin{aligned} V &= \sum_{j_1, \ldots, j_n} B^1_{j_1} \ldots B^n_{j_n} \, f^{j_1} \wedge \ldots \wedge f^{j_n} = \sum_{j_1, \ldots, j_n} B^1_{j_1} \ldots B^n_{j_n} \, \delta^{j_1 \ldots j_n}_{1 \ldots n} f^1 \wedge \ldots \wedge f^n \\ &= \det(B) \, f^1 \wedge \ldots f^n \end{aligned} \tag{3.81}$$

Let us further look at the transformation of the metric tensor g to the new basis. We have

$$g = \sum_{ij}^{n} g_{ij} \, e^i \otimes e^j = \sum_{ijkl}^{n} g_{ij} B^i_k B^j_l \, f^k \otimes f^l = \sum_{kl}^{n} g'_{kl} \, f^k \otimes f^l \tag{3.82}$$

where

$$g'_{kl} = \sum_{ij}^{n} g_{ij} B^i_k B^j_l$$

is the metric tensor in the new basis. By taking the determinant on both sides of this equation we see that

$$\det(g') = (\det(B))^2 \det(g)$$

If we denote $g = \det(g)$ (unfortunately the tensor and its determinant have the same name, but in practice it is allows clear from the context which meaning one should use) then this equation gives

$$\sqrt{|g'|} = |\det(B)| \sqrt{|g|} \tag{3.83}$$

Let us consider basis transformation that do not change the orientation of the basis, i.e. $\det(B) > 0$. Then from Eq.(3.81) and (3.83) we see that

$$\sqrt{|g|}\, e^1 \wedge \ldots \wedge e^n = \sqrt{|g|} \det(B)\, f^1 \wedge \ldots \wedge f^n = \sqrt{|g'|}\, f^1 \wedge \ldots \wedge f^n$$

Therefore the quantity

$$\Omega = \sqrt{|g|}\, e^1 \wedge \ldots \wedge e^n \tag{3.84}$$

transforms as a scalar under basis transformations. The quantity Ω is known as the *volume form*. So far we have not said anything about the properties of e^j and f^j (apart from the fact that they are a basis). If we make the special choice of an orthonormal basis $g(e_i, e_j) = \pm\delta_{ij}$ then $|g| = 1$. in terms of the corresponding dual basis e^j in V^* the volume form Ω then attains the simple form

$$\Omega = e^1 \wedge \ldots \wedge e^n$$

However, when transforming this equation to a general other basis, one should remember that there is a hidden $\sqrt{|g|} = 1$ in this equation.

3.4 The Hodge $\star$ operator

We note that the space $\Omega^p(V)$ of anti-symmetric tensors of order p on a vector space V of dimension n has the same dimension as the space $\Omega^{n-p}(V)$ of anti-symmetric tensors of order $n-p$, i.e.

$$\dim \Omega^p(V) = \binom{n}{p} = \binom{n}{n-p} = \dim \Omega^{n-p}(V).$$

This opens up the possibility of an invertible linear mapping

$$\star : \Omega^p(V) \to \Omega^{n-p}(V)$$

known as the *Hodge star* operator that assigns a tensor of order $n-p$ to a tensor of order p.
It is not difficult to interpret this geometrically. If we consider a p-dimensional plane W in a n-dimensional space, then the space $W^\perp$ spanned by the vectors orthogonal to W is $(n-p)$-dimensional. Note that to speak about orthogonality we first need to define a metric. We have seen in Eq.(3.80) that the p-form $e^{i_1} \wedge \ldots \wedge e^{i_p}$ can be regarded as spanning a p-dimensional volume. We can therefore try to construct the mapping $\star$ by assigning to a p-form that spans W a $(n-p)$-form that spans $W^\perp$. Let us draw this pictorially. To draw orthogonal axes we assume an orthogonal basis. Let us take $n = 3$ and the 2-form $e^1 \wedge e^2 \in \Omega^2$ and $\langle e^i, e^j \rangle = \pm\delta^i_j$ as in Eq.(3.66). We then have

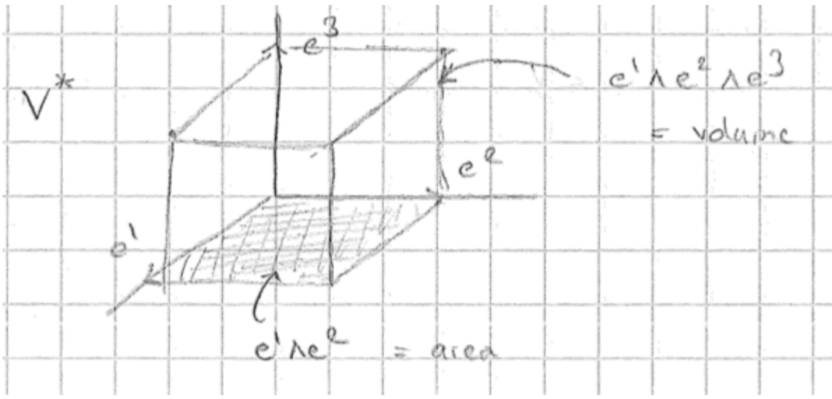

Figure 3.12: Geometrical interpretation of the wedge product

If we imagine $e^1 \wedge e^2$ to be spanning a square in the (e^1, e^2)-plane then the orthogonal space is spanned by e^3, so we would write

$$\star(e^1 \wedge e^2) = \pm e^3$$

where we would still need a good convention for the $\pm$ sign in this equation. Similarly we would have

$$\star(e^1 \wedge e^3) = \pm e^2 \qquad \star(e^2 \wedge e^3) = \pm e^1.$$

The signs can not be chosen arbitrarily if we want the mapping $\star$ to be independent of the basis. We can see this in a simpler example where V is two-dimensional and $\Omega^p = \Omega^1 = V^*$.

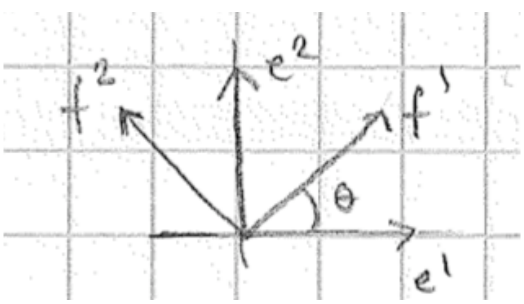

Figure 3.13: Two different orthonormal bases

Consider two orthogonal vectors e^1 and e^2 with respect to the standard Euclidean metric

$$g_{ij} = g^{ij} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

i.e. $\langle e^1, e^2 \rangle = 0$ and $\langle e^1, e^1 \rangle = \langle e^2, e^2 \rangle = 1$ (see again Eq.(3.80)). From the picture we see that

$$\begin{aligned} \star e^1 &= \alpha e^2 \\ \star e^2 &= \beta e^1 \end{aligned} \qquad (3.85)$$

where $\alpha, \beta = \pm 1$. Let us now consider a different orthonormal basis (f^1, f^2). Since the mapping $\star$ should be basis independent we want

$$\begin{aligned} \star f^1 &= \alpha f^2 \\ \star f^2 &= \beta f^1 \end{aligned} \qquad (3.86)$$

with the same values for α and β as before in Eq.(3.85). The bases (e^1, e^2) and (f^1, f^2) are related by the orthogonal transformation

$$\begin{pmatrix} f^1 \\ f^2 \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} e^1 \\ e^2 \end{pmatrix}$$

One readily checks that $\langle f^1, f^1\rangle = \langle f^2, f^2\rangle = 1$ and $\langle f^1, f^2\rangle = 0$. Since we want the operation $\star$ to be linear we must have

$$\begin{aligned} \alpha(-\sin\theta\, e^1 + \cos\theta\, e^2) &= \alpha f^2 = \star f^1 = \star(\cos\theta\, e^1 + \sin\theta\, e^2) \\ &= \cos\theta \star e^1 + \sin\theta \star e^2 = \alpha\cos\theta\, e^2 + \beta\sin\theta\, e^1 \end{aligned}$$

from which we see that $\alpha = -\beta$. The equation $\star f^2 = \beta f^1$ produces the same result. We thus find the following two choices for the signs which are valid in for any orthonormal basis:

$$\begin{aligned} \star e^1 &= e^2 \\ \star e^2 &= -e^1 \end{aligned}$$

or

$$\begin{aligned} \star e^1 &= -e^2 \\ \star e^2 &= e^1 \end{aligned}$$

We will see later that these two possible choices corresponds to different choices for the orientation of the basis.

Let us now see what happens when we change the metric. Let us instead of an Euclidean metric choose a Minkowski-type of metric

$$g_{ij} = g^{ij} = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix},$$

In this case we have $\langle e^1, e^2\rangle = 0$, $\langle e^1, e^1\rangle = -1$ and $\langle e^2, e^2\rangle = 1$. A transformation to a new orthonormal basis is given by

$$\begin{pmatrix} f^1 \\ f^2 \end{pmatrix} = \begin{pmatrix} \cosh\phi & \sinh\phi \\ \sinh\phi & \cosh\phi \end{pmatrix} \begin{pmatrix} e^1 \\ e^2 \end{pmatrix}$$

One can readily check again that $\langle f^1, f^2\rangle = 0$, $\langle f^1, f^1\rangle = -1$ and $\langle f^2, f^2\rangle = 1$. If we make the same assumptions of Eqs.(3.85) and (3.86) as before then we see that

$$\begin{aligned} \alpha(\sinh\phi\, e^1 + \cosh\phi\, e^2) &= \alpha f^2 = \star f^1 = \star(\cosh\phi\, e^1 + \sinh\phi\, e^2) \\ &= \cosh\phi \star e^1 + \sinh\phi \star e^2 = \alpha\cosh\phi\, e^2 + \beta\sinh\phi\, e^1 \end{aligned}$$

from which we see that $\alpha = \beta$. So in this case we see that the only basis-independent possibilities are

$$\begin{aligned} \star e^1 &= e^2 \\ \star e^2 &= e^1 \end{aligned} \tag{3.87}$$

or

$$\begin{aligned} \star e^1 &= -e^2 \\ \star e^2 &= -e^1. \end{aligned} \tag{3.88}$$

We conclude that the choice of the signs in the mapping $\star$ must depend on the metric g. After these preliminary insights let us now try to define the mapping $\star$ more generally.

Let $\omega, \mu \in \Omega^p(V)$, then $\star\mu \in \Omega^{n-p}(V)$ and hence

$$\omega \wedge \star\mu \in \Omega^n(V).$$

But the space $\Omega^n(V)$ is one-dimensional and the volume form Ω of Eq.(3.84) forms a basis. We can thus write

$$\omega \wedge \star\mu = c(\omega, \mu)\, \Omega \tag{3.89}$$

where $c(\omega, \mu)$ is a number depending on ω and μ. We now want to define $\star\mu$ by this equation by making a suitable choice for the function $c(\omega, \mu)$. Let us, for the moment, suppose that we would be given a such function. Then the validity of Eq.(3.89) for all possible $\omega \in \Omega^p$ would uniquely assign a $(n-p)$-form $\star\mu$ to a given p-form μ. Suppose, namely, that for a given μ and all possible ω there would be two $(n-p)$-forms ν and ν' satisfying

$$\omega \wedge \nu = c(\omega, \mu)\, \Omega \tag{3.90}$$

and

$$\omega \wedge \nu' = c(\omega, \mu)\, \Omega$$

then by subtracting we have

$$\omega \wedge (\nu - \nu') = 0 \qquad \forall\, \omega \in \Omega^p(V)$$

which implies (check for yourself) that $\nu - \nu' = 0$ and hence $\nu = \nu'$. We therefore see that Eq.(3.90) has a unique solution for ν which is determined by μ and which we denote by $\nu = \star\mu$. It only remains to specify the function $c(\omega, \mu)$. We first note that, since we want the operation $\star$ to be linear, the function $c(\omega, \mu)$ is linear in both its arguments. If $\omega = \alpha_1\omega_1 + \alpha_2\omega_2$ and $\mu = \beta_1\mu_1 + \beta_2\mu_2$ then

$$\begin{aligned} c(\omega, \mu)\Omega &= (\alpha_1\omega_1 + \alpha_2\omega_2) \wedge \star\mu = \alpha_1\omega_1 \wedge \star\mu + \alpha_2\omega_2 \wedge \star\mu \\ &= (\alpha_1 c(\omega_1, \mu) + \alpha_2 c(\omega_2, \mu))\Omega \end{aligned} \tag{3.91}$$

and

$$\begin{aligned} c(\omega, \mu)\Omega &= \omega \wedge \star(\beta_1\mu_1 + \beta_2\mu_2) = \beta_1\omega \wedge \star\mu_1 + \beta_2\omega \wedge \star\mu_2 \\ &= (\beta_1 c(\omega, \mu_1) + \beta_2 c(\omega, \mu_2))\Omega \end{aligned} \tag{3.92}$$

and we therefore conclude that

$$\begin{aligned} c(\alpha_1\omega_1 + \alpha_2\omega_2, \mu) &= \alpha_1\, c(\omega_1, \mu) + \alpha_2\, c(\omega_2, \mu) \\ c(\omega, \beta_1\mu_1 + \beta_2\mu_2) &= \beta_1\, c(\omega, \mu_1) + \beta_2\, c(\omega, \mu_2) \end{aligned}$$

and we find that the function c is linear in both arguments. Although we derived that a given μ uniquely determines $\star\mu$ there is still the possibility that two different $\mu_1, \mu_2 \in \Omega^p(V)$ map to the same $\star\mu \in \Omega^{n-p}(V)$ in which case $\star$ would not be an invertible mapping. We do not want this and we therefore need the condition that if $\star\mu_1 = \star\mu_2$ then it must follow that $\mu_1 = \mu_2$. The condition $\star\mu_1 = \star\mu_2$ gives us

$$0 = \omega \wedge \star(\mu_1 - \mu_2) = c(\omega, \mu_1 - \mu_2) \quad \forall\, \omega \in \Omega^p(V)$$

If we want to conclude from this equation that $\mu_1 = \mu_2$ we need that the function c satisfies the condition

$$c(\omega, \mu) = 0 \quad \forall\, \omega \in \Omega^p(V) \quad \Rightarrow \quad \mu = 0$$

This condition together with the linearity in both arguments suggests that for c we should take a non-degenerate inner product on $\Omega^p(V)$. We learned before that the space $\Omega^{n-p}(V)$ has such an inner product given by Eq.(3.63). We therefore define

$$c(\omega, \mu) = \langle \omega, \mu \rangle$$

and the defining equation for $\star\mu$ therefore becomes

$$\omega \wedge \star\mu = \langle \omega, \mu \rangle \, \Omega \qquad \forall \, \omega \in \Omega^p(V) \tag{3.93}$$

Since we nowhere used an explicit basis this equation defines $\star\mu$ in a completely basis independent manner.

Let us now give some examples. Let us take a two-dimensional space with a Minkowski-type metric

$$g_{ij} = g^{ij} = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Then $\langle e^1, e^1 \rangle = g^{11} = -1$ and $\langle e^2, e^2 \rangle = g^{22} = 1$. Further, since $|g| = 1$ we see from Eq.(3.84) that

$$\Omega = e^1 \wedge e^2.$$

Then Eq.(3.93) tells us that

$$\begin{aligned} e^1 \wedge \star e^1 &= \langle e^1, e^1 \rangle \, e^1 \wedge e^2 = -e^1 \wedge e^2 \\ e^2 \wedge \star e^2 &= \langle e^2, e^2 \rangle \, e^1 \wedge e^2 = e^1 \wedge e^2 \end{aligned}$$

from which we deduce

$$\begin{aligned} \star e^1 &= -e^2 \\ \star e^2 &= -e^1 \end{aligned}$$

This corresponds to Eq.(3.88). If we had defined the volume form to be $\Omega = e^2 \wedge e^1$, which would correspond to a change of the orientation of the basis to (e^2, e^1) we would have obtained Eq.(3.87). We therefore see that, apart from a dependence on a metric, the Hodge star operator also depends on the chosen orientation of the basis used in the volume form. For a general one-form $\omega = w_1 e^1 + w_2 e^2$ we find that

$$\star\omega = w_1 \star e^1 + w_2 \star e^2 = -w_1 \, e^1 - w_2 \, e^2$$

Let us now go back to our first picture and take a three-dimensional Euclidean space with a metric

$$g = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

such that $\langle e^i, e^j \rangle = \delta_i^j$. Then Eq.(3.93) tells us that

$$\begin{aligned} e^1 \wedge e^2 \wedge \star(e^1 \wedge e^2) &= \langle e^1 \wedge e^2, e^1 \wedge e^2 \rangle \, e^1 \wedge e^2 \wedge e^3 = e^1 \wedge e^2 \wedge e^3 \\ e^1 \wedge e^3 \wedge \star(e^1 \wedge e^3) &= \langle e^1 \wedge e^3, e^1 \wedge e^3 \rangle \, e^1 \wedge e^2 \wedge e^3 = e^1 \wedge e^2 \wedge e^3 \\ e^2 \wedge e^3 \wedge \star(e^2 \wedge e^3) &= \langle e^2 \wedge e^3, e^2 \wedge e^3 \rangle \, e^1 \wedge e^2 \wedge e^3 = e^1 \wedge e^2 \wedge e^3 \end{aligned}$$

and we therefore find that

$$\star(e^1 \wedge e^2) = e^3 \,, \quad \star(e^1 \wedge e^3) = -e^2 \,, \quad \star(e^2 \wedge e^3) = e^1 \tag{3.94}$$

For a general tensor $T \in \Omega^2(V)$ of the form

$$T = \alpha\, e^1 \wedge e^2 + \beta\, e^3 \wedge e^1 + \gamma\, e^2 \wedge e^3$$

we thus find

$$\star T = \gamma\, e^1 + \beta\, e^2 + \alpha\, e^3$$

After these examples it is not difficult to derive the action of the Hodge star operator on a general p-form in an orthonormal basis. We have

$$\begin{aligned} e^{i_1} \wedge \ldots \wedge e^{i_p} \wedge \star(e^{i_1} \wedge \ldots \wedge e^{i_p}) &= \langle e^{i_1} \wedge \ldots \wedge e^{i_p}, e^{i_1} \wedge \ldots \wedge e^{i_p} \rangle\, e^1 \wedge \ldots \wedge e^n \\ &= g^{i_1 i_1} \ldots g^{i_p i_p}\, e^1 \wedge \ldots \wedge e^n \end{aligned}$$

where we used Eq.(3.66). From this equation we find that

$$\star(e^{i_1} \wedge \ldots \wedge e^{i_p}) = \epsilon_{i_1 \ldots i_n}\, g^{i_1 i_1} \ldots g^{i_p i_p}\, e^{i_{p+1}} \wedge \ldots \wedge e^{i_n} \tag{3.95}$$

where $(i_{p+1}, \ldots, i_n)$ are the numbers $(1, \ldots, n)$ with the numbers $(i_1, \ldots, i_p)$ removed (the order of the remaining numbers is irrelevant since both the wedge product and the ϵ-tensor are anti-symmetric) . For instance, for a 3-form defined on a 5-dimensional space we have

$$\star(e^1 \wedge e^3 \wedge e^5) = \epsilon_{13524}\, g^{11} g^{33} g^{55}\, e^2 \wedge e^4 = -g^{11} g^{33} g^{55}\, e^2 \wedge e^4 \tag{3.96}$$

We can also calculate the action of the Hodge star in a general non-orthonormal basis using Eq.(3.65). In that case we have

$$\begin{aligned} e^{i_1} \wedge \ldots \wedge e^{i_p} \wedge \star(e^{i_1} \wedge \ldots \wedge e^{i_p}) &= \langle e^{i_1} \wedge \ldots \wedge e^{i_p}, e^{i_1} \wedge \ldots \wedge e^{i_p} \rangle \sqrt{|g|}\, e^1 \wedge \ldots \wedge e^n \\ &= \det(G) \sqrt{|g|}\, e^1 \wedge \ldots \wedge e^n \end{aligned}$$

where G is the $p \times p$ matrix with entries $G_{kl} = g^{i_k i_l}$ for $k, l = 1, \ldots, p$. The equivalent of Eq.(3.95) in a non-orthonormal basis therefore becomes

$$\star(e^{i_1} \wedge \ldots \wedge e^{i_p}) = \epsilon_{i_1 \ldots i_n}\, \det(G) \sqrt{|g|}\, e^{i_{p+1}} \wedge \ldots \wedge e^{i_n} \tag{3.97}$$

The equivalent of example (3.95) in a general basis is then

$$\star(e^1 \wedge e^3 \wedge e^5) = \epsilon_{13524} \begin{vmatrix} g^{11} & g^{13} & g^{15} \\ g^{31} & g^{33} & g^{35} \\ g^{51} & g^{53} & g^{55} \end{vmatrix} \sqrt{|g|}\, e^2 \wedge e^4 \tag{3.98}$$

which clearly reduces to Eq.(3.95) for an orthonormal basis. Eq.(3.97) is sometimes also written in a different way as

$$\star(e^{i_1} \wedge \ldots \wedge e^{i_p}) = \frac{1}{(n-p)!} \sum_{i_{p+1} \ldots i_n}^{n} \epsilon_{i_1 \ldots i_n}\, \det(G) \sqrt{|g|}\, e^{i_{p+1}} \wedge \ldots \wedge e^{i_n}$$

which follows from the fact that both the ϵ-tensor and the wedge product are anti-symmetric. We can further write out $\det(G)$ explicitly as

$$\det(G) = \sum_{j_1 \ldots j_p}^{n} \delta^{i_1 \ldots i_p}_{j_1 \ldots j_p}\, g^{i_1 j_1} \ldots g^{i_p j_p}$$

Inserting this expression then yields

$$\begin{aligned}\star(e^{i_1}\wedge\ldots\wedge e^{i_p}) &= \frac{1}{(n-p)!}\sum_{j_1\ldots j_p,i_{p+1}\ldots i_n}^{n} \epsilon_{i_1\ldots i_n}\delta^{i_1\ldots i_p}_{j_1\ldots j_p}\, g^{i_1j_1}\ldots g^{i_pj_p}\sqrt{|g|}\, e^{i_{p+1}}\wedge\ldots\wedge e^{i_n}\\ &= \frac{1}{(n-p)!}\sum_{j_1\ldots j_p,i_{p+1}\ldots i_n}^{n} \epsilon_{j_1\ldots j_p,i_{p+1}\ldots i_n}\, g^{i_1j_1}\ldots g^{i_pj_p}\sqrt{|g|}\, e^{i_{p+1}}\wedge\ldots\wedge e^{i_n}\end{aligned} \tag{3.99}$$

where we used that $\epsilon_{i_1\ldots i_n}\delta^{i_1\ldots i_p}_{j_1\ldots j_p}=\epsilon_{j_1\ldots j_p,i_{p+1}\ldots i_n}$. We will use this expression to calculate the components of $\star\omega$ for a general p-form ω

$$\omega=\sum_{i_1\ldots i_p}^{n}\omega_{i_1\ldots i_p}e^{i_1}\otimes\ldots\otimes e^{i_p}=\sum_{i_1<\ldots<i_p}^{n}\omega_{i_1\ldots i_p}e^{i_1}\wedge\ldots\wedge e^{i_p} \tag{3.100}$$

Since the components $\omega_{i_1\ldots i_p}$ are anti-symmetric under interchange of indices we can equivalently write

$$\omega=\frac{1}{p!}\sum_{i_1\ldots i_p}^{n}\omega_{i_1\ldots i_p}e^{i_1}\wedge\ldots\wedge e^{i_p}$$

From Eq.(3.99) we then obtain that

$$\begin{aligned}\star\omega &= \frac{1}{p!}\sum_{i_1\ldots i_p}^{n}\omega_{i_1\ldots i_p}\star(e^{i_1}\wedge\ldots\wedge e^{i_p})\\ &= \frac{1}{p!(n-p)!}\sum_{j_1\ldots j_p,i_1\ldots i_n}^{n}\epsilon_{j_1\ldots j_p,i_{p+1}\ldots i_n}\omega_{i_1\ldots i_p}\, g^{i_1j_1}\ldots g^{i_pj_p}\sqrt{|g|}\, e^{i_{p+1}}\wedge\ldots\wedge e^{i_n}\\ &= \frac{1}{p!(n-p)!}\sum_{j_1\ldots j_p,i_{p+1}\ldots i_n}^{n}\omega^{j_1\ldots j_p}\epsilon_{j_1\ldots j_p,i_{p+1}\ldots i_n}\sqrt{|g|}\, e^{i_{p+1}}\wedge\ldots\wedge e^{i_n}\\ &= \frac{1}{(n-p)!}\sum_{i_{p+1}\ldots i_n}^{n}(\star\omega)_{i_{p+1}\ldots i_n}\, e^{i_{p+1}}\wedge\ldots\wedge e^{i_n}\\ &= \sum_{i_{p+1}<\ldots<i_n}^{n}(\star\omega)_{i_{p+1}\ldots i_n}\, e^{i_{p+1}}\wedge\ldots\wedge e^{i_n}\end{aligned}$$

where we defined

$$\begin{aligned}(\star\omega)_{i_{p+1}\ldots i_n} &= \frac{1}{p!}\sqrt{|g|}\sum_{j_1\ldots j_p}^{n}\omega^{j_1\ldots j_p}\epsilon_{j_1\ldots j_p,i_{p+1}\ldots i_n}\\ &= \sqrt{|g|}\sum_{j_1<\ldots<j_p}^{n}\omega^{j_1\ldots j_p}\epsilon_{j_1\ldots j_p,i_{p+1}\ldots i_n}\end{aligned} \tag{3.101}$$

where we raised the indices on ω. This equation gives the desired explicit expression for the components of the $(n-p)$-form $\star\omega$ in terms of the components of the p-form ω.

Let us now apply this to the important case of the electromagnetic field tensor about which we will hear much more later. For the moment it is sufficient to know that it is a 2-form on a

4-dimensional vector space with Minkowski metric. We write it as

$$
\begin{aligned}
F &= \sum_{\mu\nu}^{4} F_{\mu\nu}\, e^{\mu} \otimes e^{\nu} = \sum_{\mu<\nu} F_{\mu\nu}\, (e^{\mu} \otimes e^{\nu} - e^{\nu} \otimes e^{\mu}) = \sum_{\mu<\nu} F_{\mu\nu}\, e^{\mu} \wedge e^{\nu} \\
&= \frac{1}{2} \sum_{\mu\nu}^{4} F_{\mu\nu}\, e^{\mu} \wedge e^{\nu}
\end{aligned}
$$

where

$$
F_{\mu\nu} = \begin{pmatrix} 0 & -E_{\mathrm{x}} & -E_{\mathrm{y}} & -E_{\mathrm{z}} \\ E_{\mathrm{x}} & 0 & B_{\mathrm{z}} & -B_{\mathrm{y}} \\ E_{\mathrm{y}} & -B_{\mathrm{z}} & 0 & B_{\mathrm{x}} \\ E_{\mathrm{z}} & B_{\mathrm{y}} & -B_{\mathrm{x}} & 0 \end{pmatrix} \tag{3.102}
$$

We see that $F_{\mu\nu} = -F_{\nu\mu}$ is an anti-symmetric and covariant second order tensor. The components E_j and B_j have the physical meaning of electric and magnetic field components. To use Eqn.(3.101) we need to raise the indices on $F_{\mu\nu}$. We do this with the Minkowski metric

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

This gives

$$
F^{\mu\nu} = \sum_{\tau\rho}^{4} g^{\mu\tau} g^{\nu\rho} F_{\tau\rho} = g^{\mu\mu} g^{\nu\nu} F_{\mu\nu}
$$

since the metric tensor is diagonal. This gives

$$
F^{\mu\nu} = \begin{pmatrix} 0 & E_{\mathrm{x}} & E_{\mathrm{y}} & E_{\mathrm{z}} \\ -E_{\mathrm{x}} & 0 & B_{\mathrm{z}} & -B_{\mathrm{y}} \\ -E_{\mathrm{y}} & -B_{\mathrm{z}} & 0 & B_{\mathrm{x}} \\ -E_{\mathrm{z}} & B_{\mathrm{y}} & -B_{\mathrm{x}} & 0 \end{pmatrix} \tag{3.103}
$$

Now we can calculate $(\star F)_{\mu\nu}$ from Eq.(3.101). Since in our case $|g| = 1$ we find

$$
(\star F)_{\mu\nu} = \sum_{\alpha<\beta} F^{\alpha\beta} \epsilon_{\alpha\beta\mu\nu} = \sum_{\alpha<\beta} \epsilon_{\mu\nu\alpha\beta} F^{\alpha\beta}
$$

More explicitly this gives

$$
\begin{aligned}
(\star F)_{12} &= \epsilon_{1234} F^{34} = B_{\mathrm{x}} & \qquad (\star F)_{13} &= \epsilon_{1324} F^{24} = -(-B_{\mathrm{y}}) = B_{\mathrm{y}} \\
(\star F)_{14} &= \epsilon_{1423} F^{23} = B_{\mathrm{z}} & \qquad (\star F)_{23} &= \epsilon_{2314} F^{14} = E_{\mathrm{z}} \\
(\star F)_{24} &= \epsilon_{2413} F^{13} = -E_{\mathrm{y}} & \qquad (\star F)_{34} &= \epsilon_{3412} F^{12} = E_{\mathrm{x}}
\end{aligned}
$$

and we therefore find that

$$
(\star F)_{\mu\nu} = \begin{pmatrix} 0 & B_{\mathrm{x}} & B_{\mathrm{y}} & B_{\mathrm{z}} \\ -B_{\mathrm{x}} & 0 & E_{\mathrm{z}} & -E_{\mathrm{y}} \\ -B_{\mathrm{y}} & -E_{\mathrm{z}} & 0 & E_{\mathrm{x}} \\ -B_{\mathrm{z}} & E_{\mathrm{y}} & -E_{\mathrm{x}} & 0 \end{pmatrix} \tag{3.104}
$$

If we compare this equation to Eq.(3.102) we see that making the Hodge star operation amounts to the operation

$$
E_i \to -B_i \quad , \quad B_i \to E_i
$$

on the electric and magnetic fields. Looking at the Maxwell equations we see that the Hodge star transforms one set of Maxwell's equations, namely

$$\nabla \cdot \mathbf{B} = 0 \quad , \quad \nabla \times \mathbf{E} + \frac{1}{c}\frac{\partial \mathbf{B}}{\partial t} = 0$$

into another pair

$$\nabla \cdot \mathbf{E} = 0 \quad , \quad \nabla \times \mathbf{B} - \frac{1}{c}\frac{\partial \mathbf{E}}{\partial t} = 0$$

(in the absence of charges and currents). This will be explained more fully later.

Let us finalize our discussion by deriving some further properties of the Hodge star operator. The first one that we will prove is

$$\langle \star\alpha, \star\beta \rangle = \langle \alpha, \beta \rangle \, \mathrm{sign}(g) \tag{3.105}$$

where $\alpha, \beta \in \Omega^p(V)$ and $\mathrm{sign}(g)$ is the sign of the determinant of the metric (e.g. -1 for the Minkowski metric). This can be shown by a more or less straightforward calculation. According to the definition of the inner product we have

$$\langle \star\alpha, \star\beta \rangle = \frac{1}{(n-p)!} \sum_{i_{p+1}\dots i_n}^{n} (\star\alpha)_{i_{p+1}\dots i_n} (\star\beta)^{i_{p+1}\dots i_n}$$

Let us first raise the indices of $\star\beta$. According to Eq.(3.101) we have

$$\begin{aligned}
(\star\beta)^{i_{p+1}\dots i_n} &= \sum_{j_{p+1}\dots j_n}^{n} (\star\beta)_{j_{p+1}\dots j_n} \, g^{i_{p+1}j_{p+1}} \dots g^{i_n j_n} \\
&= \frac{1}{p!}\sqrt{|g|} \sum_{j_1\dots j_n}^{n} \beta^{j_1\dots j_p} \epsilon_{j_1\dots j_n} \, g^{i_{p+1}j_{p+1}} \dots g^{i_n j_n} \\
&= \frac{1}{p!}\sqrt{|g|} \sum_{j_1\dots j_n, i_1\dots i_p}^{n} \beta_{i_1\dots i_p} \epsilon_{j_1\dots j_n} \, g^{i_1 j_1} \dots g^{i_n j_n}
\end{aligned} \tag{3.106}$$

Now we note that

$$\epsilon^{i_1\dots i_n} = \sum_{j_1\dots j_n}^{n} \epsilon_{j_1\dots j_n} \, g^{i_1 j_1} \dots g^{i_n j_n} = \epsilon_{i_1\dots i_n} \epsilon^{1\dots n}$$

since $\epsilon^{i_1\dots i_n}$ is anti-symmetric and we therefore need only the component $\epsilon^{1\dots n}$ to determine all other components. The component $\epsilon^{1\dots n}$ is simply given by

$$\epsilon^{1\dots n} = \sum_{j_1\dots j_n}^{n} \epsilon_{j_1\dots j_n} \, g^{1 j_1} \dots g^{n j_n} = \det(g^{ij}) = \det(g_{ij})^{-1} = \frac{1}{g} \tag{3.107}$$

We therefore find that

$$(\star\beta)^{i_{p+1}\dots i_n} = \frac{1}{p!} \frac{\mathrm{sign}(g)}{\sqrt{|g|}} \sum_{i_1\dots i_p}^{n} \beta_{i_1\dots i_p} \epsilon_{i_1\dots i_n}$$

With this equation and the explicit form of $\star\alpha$ from Eq.(3.101) we then find that

$$
\begin{aligned}
\langle \star\alpha, \star\beta \rangle &= \frac{\mathrm{sign}(g)}{p!p!(n-p)!} \sum_{j_1\dots j_p, i_1\dots i_n}^{n} \alpha^{j_1\dots j_p} \epsilon_{j_1\dots j_p, i_{p+1}\dots i_n} \epsilon_{i_1\dots i_n} \beta_{i_1\dots i_p} \\
&= \frac{\mathrm{sign}(g)}{p!p!} \sum_{j_1\dots j_p, i_1\dots i_p}^{n} \alpha^{j_1\dots j_p} \delta^{i_1\dots i_p}_{j_1\dots j_p} \beta_{i_1\dots i_p} \\
&= \mathrm{sign}(g) \frac{1}{p!} \sum_{j_1\dots j_p}^{n} \alpha^{j_1\dots j_p} \beta_{j_1\dots j_p} = \mathrm{sign}(g) \langle \alpha, \beta \rangle
\end{aligned}
$$

which proves Eq.(3.105). This equation is very useful for moving around stars in equations. For instance, we can deduce that

$$
\begin{aligned}
\mathrm{sign}(g) \star \alpha \wedge \star\star\, \beta &= \mathrm{sign}(g) \langle \star\alpha, \star\beta \rangle \Omega = \langle \alpha, \beta \rangle \Omega = \langle \beta, \alpha \rangle \Omega \\
&= \beta \wedge \star\alpha = (-1)^{p(n-p)} \star \alpha \wedge \beta
\end{aligned}
$$

where in the last step we used the property (3.75) of the wedge product. From this equation we therefore deduce that for a p-form β on a n-dimensional space with metric g we have

$$
\star\star\, \beta = (-1)^{p(n-p)}\, \mathrm{sign}(g)\, \beta \tag{3.108}
$$

For instance, for the electromagnetic field tensor of Eq.(3.102) we have

$$
(\star\star F)_{\mu\nu} = (-1)^{2(4-2)} (-1) F_{\mu\nu} = -F_{\mu\nu}
$$

With Eq.(3.108) we can also write Eq.(3.93) in a different way. If we let the p-form μ in Eq.(3.93) be equal to $\star\nu$ (so ν is an $(n-p)$-form) then we can write

$$
\omega \wedge \star\star\, \nu = \langle \omega, \star\nu \rangle \Omega
$$

which, with Eq.(3.108) gives

$$
(-1)^{p(n-p)}\, \mathrm{sign}(g)\, \omega \wedge \nu = \langle \omega, \star\nu \rangle \Omega \tag{3.109}
$$

Let now ω be a 1-form and take

$$
\nu = \nu_1 \wedge \ldots \wedge \nu_{n-1}
$$

where $\nu_1, \ldots, \nu_{n-1}$ are also 1-forms, i.e. elements of V^*. Then Eq.(3.109) tells that

$$
(-1)^{n(n-1)} \mathrm{sign}(g)\, \omega \wedge \nu_1 \wedge \ldots \nu_{n-1} = \langle \omega, \star(\nu_1 \wedge \ldots \wedge \nu_{n-1}) \rangle \Omega
$$

Now the left-hand side of this equation vanishes when ω is equal to one of the ν_j for $j = 1, \ldots, n-1$ or a linear combination of them. We therefore see that

$$
\langle \nu_j, \star(\nu_1 \wedge \ldots \wedge \nu_{n-1}) \rangle = 0 \qquad j = 1, \ldots, n-1 \tag{3.110}
$$

The dual vector $\star(\nu_1 \wedge \ldots \wedge \nu_{n-1})$ is therefore orthogonal to all the vectors $\nu_1, \ldots, \nu_{n-1}$. This represents the generalization of the outer product in three-dimensional spaces to vector spaces with arbitrary dimensions and metrics. For the familiar case of a three-dimensional vector space with Euclidean metric we have that if

$$
\begin{aligned}
a &= a_1 e^1 + a_2 e^2 + a_3 e^3 \\
b &= b_1 e^1 + b_2 e^2 + b_3 e^3
\end{aligned}
$$

then

$$\begin{aligned} a \wedge b &= (a_1 e^1 + a_2 e^2 + a_3 e^3) \wedge (b_1 e^1 + b_2 e^2 + b_3 e^3) \\ &= (a_1 b_2 - a_2 b_1) e^1 \wedge e^2 + (a_1 b_3 - a_3 b_1) e^1 \wedge e^3 + (a_2 b_3 - a_3 b_2) e^2 \wedge e^3 \end{aligned}$$

From Eq.(3.94) we therefore find that

$$\begin{aligned} \star(a \wedge b) &= (a_2 b_3 - a_3 b_2) e^1 - (a_1 b_3 - a_3 b_1) e^2 + (a_1 b_2 - a_2 b_1) e^3 \\ &= \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \times \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} \end{aligned}$$

i.e. we recover the familiar outer product in three dimensions.

3.5 Tensors along curves and surfaces

In the previous sections we discussed the properties of vectors and tensors. In this section we will discuss how tensors can be applied applied to tangent vectors to curves and surfaces. The only thing we need to do for this is to replace the general vector space V used in the previoussections by the tangent space T_pM at the point p of some manifold M. Similarly, we need to replace the dual space V^* by T_pM^*.
Let us illustrate everything with an example. Since we already used the sphere several times, we will use the torus instead.

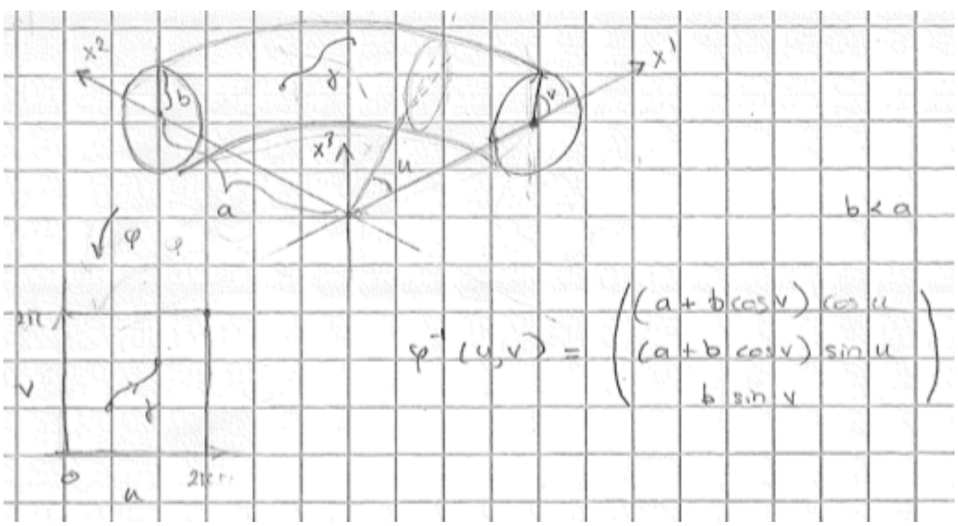

Figure 3.14: Parametrization of the torus.

As a coordinate map we use

$$\varphi^{-1}(u, v) = \begin{pmatrix} (a + b\cos v)\cos u \\ (a + b\cos v)\sin u \\ b\sin v \end{pmatrix}$$

Here $\varphi : M \to \mathbb{R}^2$ is a coordinate map from the torus to the two-dimensional space $\mathbb{R}^2$. A curve

$$\gamma(t) = (u(t), v(t))$$

parametrized by t (which may be interpreted as time, for instance when we imagine the motion of a particle along the surface) represents via the inverse coordinate map $\varphi^{-1}(u(t), v(t))$ a curve on the torus. Let $f : M \to \mathbb{R}$ be a scalar function defined on M, such as a static temperature field. We can calculate the rate of change of change of the temperature along $\gamma(t)$ by calculating the derivative

$$\frac{\partial f(\gamma(t))}{\partial t} = \frac{\partial f(u(t), v(t))}{\partial t} = \frac{\partial u}{\partial t}\frac{\partial f}{\partial u} + \frac{\partial v}{\partial t}\frac{\partial f}{\partial v}$$

We write this as

$$\frac{\partial \gamma}{\partial t}(f) = \left[\frac{\partial u}{\partial t}\frac{\partial}{\partial u} + \frac{\partial v}{\partial t}\frac{\partial}{\partial v}\right](f)$$

From our discussion in the previous Chapter we see that

$$\frac{\partial \gamma}{\partial t} = \frac{\partial u}{\partial t}\frac{\partial}{\partial u} + \frac{\partial v}{\partial t}\frac{\partial}{\partial v} = (\frac{\partial u}{\partial t}, \frac{\partial v}{\partial t})$$

is a vector in point $\gamma(t)$. We may call it the velocity vector at point $\gamma(t)$. The most general vector field on the torus is of the form

$$\nu = a(u, v)\frac{\partial}{\partial u} + b(u, v)\frac{\partial}{\partial v} \tag{3.111}$$

Any vector field on the torus can be completely described by such two-dimensional vectors. if we want to map these vectors to three-component vectors ω in the surrounding space we have to push forward the vectors Eq.(3.111) with the embedding map $i : M \to \mathbb{R}^3$ given by

$$\begin{aligned} x^1(u, v) &= (a + b\cos v)\cos u \\ x^2(u, v) &= (a + b\cos v)\sin u \\ x^3(u, v) &= b\sin v \end{aligned} \tag{3.112}$$

We can then let ν of Eq.(3.111) act on any function $f(x^1, x^2, x^3)$ defined on $\mathbb{R}^3$ by

$$\begin{aligned} \omega(f) &= i_*\nu(f) = \nu(f \circ i) \\ &= \left(a(u, v)\frac{\partial}{\partial u} + b(u, v)\frac{\partial}{\partial v}\right) f(x^1(u, v), x^2(u, v), x^3(u, v)) \\ &= \sum_{j=1}^{3}\left(a\frac{\partial x^j}{\partial u} + b\frac{\partial x^j}{\partial v}\right)\frac{\partial f}{\partial x^j} = \sum_{j=1}^{3}\omega^j\frac{\partial f}{\partial x^j} \end{aligned} \tag{3.113}$$

where

$$i_*\nu = \begin{pmatrix} \omega^1 \\ \omega^2 \\ \omega^3 \end{pmatrix} = \begin{pmatrix} \frac{\partial x^1}{\partial u} & \frac{\partial x^1}{\partial v} \\ \frac{\partial x^2}{\partial u} & \frac{\partial x^2}{\partial v} \\ \frac{\partial x^3}{\partial u} & \frac{\partial x^3}{\partial v} \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix} \tag{3.114}$$

In particular we have

$$i_*(\frac{\partial}{\partial u}) = \begin{pmatrix} \frac{\partial x^1}{\partial u} & \frac{\partial x^1}{\partial v} \\ \frac{\partial x^2}{\partial u} & \frac{\partial x^2}{\partial v} \\ \frac{\partial x^3}{\partial u} & \frac{\partial x^3}{\partial v} \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{\partial x^1}{\partial u} \\ \frac{\partial x^2}{\partial u} \\ \frac{\partial x^3}{\partial u} \end{pmatrix}$$

Let us denote this vector by $\partial\mathbf{x}/\partial u$. Then for the case of the torus we have

$$i_*(\frac{\partial}{\partial u}) = \frac{\partial \mathbf{x}}{\partial u} = \begin{pmatrix} -(a + b\cos v)\sin u \\ (a + b\cos v)\cos u \\ 0 \end{pmatrix} \tag{3.115}$$

which describes the tangent vector to the torus in the u-direction. Similarly we have

$$i_*(\frac{\partial}{\partial v}) = \frac{\partial \mathbf{x}}{\partial v} = \begin{pmatrix} -b\sin v \cos u \\ -b \sin v \sin u \\ b\cos v \end{pmatrix} \tag{3.116}$$

So far the discussion of tangent vectors. Let us now discuss the cotangent vectors. We first discuss the general case and come back to the example of the torus.
A covariance tensor of order 1, or simply a covector on a tangent space T_pM is a linear mapping from T_pM to the real numbers. One such a mapping is suggested by the definition of the tangent vector itself as a linear operation on functions f on an manifold. We define a mapping $df : T_pM \to \mathbb{R}$ by

$$df(v) = v(f) \tag{3.117}$$

for a given choice of function f. The mapping df is obviously linear. For if $v, w \in T_pM$ and $\alpha, \beta \in \mathbb{R}$ then we have

$$df(\alpha\, v + \beta\, w) = (\alpha\, v + \beta\, w)(f) = \alpha\, v(f) + \beta\, w(f) = \alpha\, df(v) + \beta\, df(w)$$

Let us write out Eq.(3.117) in a coordinate system $x = (x^1, \ldots, x^n)$. We have

$$df(v) = v(f) = \sum_{j=1}^{n} v^j(x) \frac{\partial f}{\partial x^j} \tag{3.118}$$

In particular, if we choose $f(x) = x^k$, i.e. just the coordinate function, we obtain

$$dx^k(v) = \sum_{j=1}^{n} v^j(x) \frac{\partial x^k}{\partial x^j} = \sum_{j=1}^{n} v^j(x) \delta^k_j = v^k(x)$$

If we now take

$$v = \frac{\partial}{\partial x^j} = (0, \ldots, 0, 1, 0 \ldots, 0)$$

where there is a 1 on the j-th position, then

$$dx^k(\frac{\partial}{\partial x^j}) = \delta^k_j \tag{3.119}$$

So the covectors form a basis which is dual to the basis vectors $\partial/\partial x^j$ of T_pM. Let us now take an arbitrary covector $\omega \in T_pM^*$ then

$$\omega(v) = \omega(\sum_{j=1}^{n} v^j \frac{\partial}{\partial x^j}) = \sum_{j=1}^{n} v^j \omega(\frac{\partial}{\partial x^j}) = \sum_{j=1}^{n} v^j \omega_j = \sum_{j=1}^{n} \omega_j\, dx^j(v) \tag{3.120}$$

where we defined

$$\omega_j = \omega(\frac{\partial}{\partial x^j})$$

We thus have

$$\omega = \sum_{j=1}^{n} \omega_j\, dx^j = (\omega_1, \ldots, \omega_n) \tag{3.121}$$

and we find that any dual vector ω in T_pM^* can be expressed in terms of the dual basis dx^j. In particular for the covector df of Eq.(3.118) we have

$$df = \sum_{j=1}^{n} \frac{\partial f}{\partial x^j}\, dx^j = (\frac{\partial f}{\partial x^1}, \ldots, \frac{\partial f}{\partial x^n}) \tag{3.122}$$

which is just the familiar gradient of the function f. Let us now see how the covectors transform under coordinate transformations. Taking in Eq.(3.122) the function $f(x) = x'^k(x)$ for a coordinate function of a new coordinate system we have

$$dx'^k = \sum_{j=1}^{n} \frac{\partial x'^k}{\partial x^j}\, dx^j \tag{3.123}$$

If we therefore write ω of Eq.(3.121) in two different coordinate systems

$$\omega = \sum_{j=1}^{n} \omega_j\, dx^j = \sum_{j=1}^{n} \omega'_j\, dx'^j$$

then we see from Eq.(3.123) that

$$\sum_{j=1}^{n} \omega'_j\, dx'^j = \sum_{j,k=1}^{n} \omega'_j \frac{\partial x'^j}{\partial x^k} dx^k = \sum_{k=1}^{n} \omega_k\, dx^k$$

where

$$\omega_k = \sum_{j=1}^{n} \frac{\partial x'^j}{\partial x^k}\, \omega'_j \tag{3.124}$$

This should be compared to the transformation law for vectors.

$$v = \sum_{j=1}^{n} v'^j \frac{\partial}{\partial x'^j} = \sum_{j,k=1}^{n} v'^j \frac{\partial x^k}{\partial x'^j} \frac{\partial}{\partial x^k} = \sum_{k=1}^{n} v^k \frac{\partial}{\partial x^k}$$

i.e.

$$v^k = \sum_{j=1}^{n} \frac{\partial x^k}{\partial x' j}\, v'^j \tag{3.125}$$

We see that vectors and convectors transform opposite to each other under coordinate transformations. A commonly occurring covector in physics is the force vector. If we write Newton's equations for the force in components

$$F_j = -\frac{\partial V}{\partial x^j} \tag{3.126}$$

where $V(x^1, \ldots, x^n)$ is a potential function. Then we see that under coordinate transformations to new coordinates $(x'^1, \ldots, x'^n)$

$$F_j = \sum_{k=1}^{n} -\frac{\partial V}{\partial x'^k} \frac{\partial x'^k}{\partial x^j} = \sum_{k=1^n} \frac{\partial x'^k}{\partial x^j}\, F'_k \tag{3.127}$$

and hence F transforms as a covector. So if we write in three dimensions

$$F = (F_1, F_2, F_3) = F_1 dx^1 + F_2 dx^2 + F_3 dx^3$$

then Eq.(3.126) can be written as

$$F = -dV$$

How this fits with $\mathbf{F} = m d\mathbf{v}/dt$ where we are used to call both sides vectors, we will see later. If we act with the force covector F on the velocity vector

$$v = v^1(t) \frac{\partial}{\partial x^1} + v^2(t) \frac{\partial}{\partial x^2} + v^3(t) \frac{\partial}{\partial x^3}$$

we have

$$F(v) = \sum_{k=1}^{n} F_k dx^k (\sum_{j=1}^{n} v^j \frac{\partial}{\partial x^j}) = \sum_{j,k=1}^{n} F_k v^j dx^k (\frac{\partial}{\partial x^j}) = \sum_{j,k=1}^{n} F_k v^j \delta_j^k = \sum_{j=1}^{n} F_j v^j = \mathbf{F} \cdot \mathbf{v}$$

This is a coordinate invariant scalar (under spatial transformations) representing the work done by the force $\mathbf{F}$ on a particle moving at velocity $\mathbf{v}$.

Now that we have defined the tangent space $V = T_pM$ with basis $e_j = \partial/\partial x^j$ and the dual tangent space $V^* = T_pM^*$ with basis dx^j we can directly copy all results of the previous Chapter and define tensors. A k-th order covariant tensor is a multilinear mapping

$$T : T_pM \times \ldots \times T_pM \to \mathbb{R}$$

In the basis $\{dx^k\}$ of T_pM^* it has the form

$$T = \sum_{j_1 \ldots j_k}^{n} T_{j_1 \ldots j_k}(x)\, dx^{j_1} \otimes \ldots \otimes dx^{j_k} \tag{3.128}$$

From Eq.(3.123) we can the readily derive the transformation law for the components of T. We have

$$\begin{aligned} T &= \sum_{j_1 \ldots j_p}^{n} T'_{j_1 \ldots j_p}\, dx'^{j_1} \otimes \ldots \otimes dx'^{j_p} \\ &= \sum_{j_1 \ldots j_p, k_1 \ldots k_p}^{n} T'_{j_1 \ldots j_p} \frac{\partial x'^{j_1}}{\partial x^{k_1}} \cdots \frac{\partial x'^{j_p}}{\partial x^{k_p}}\, dx^{k_1} \otimes \ldots \otimes dx^{k_p} \\ &= \sum_{k_1 \ldots k_p}^{n} T_{k_1 \ldots k_p}\, dx^{k_1} \otimes \ldots \otimes dx^{k_p} \end{aligned} \tag{3.129}$$

and hence

$$T_{k_1 \ldots k_p} = \sum_{j_1 \ldots j_p}^{n} T'_{j_1 \ldots j_p} \frac{\partial x'^{j_1}}{\partial x^{k_1}} \cdots \frac{\partial x'^{j_p}}{\partial x^{k_p}} \tag{3.130}$$

which is generalization of Eq.(3.124). Similarly we can also consider contravariant tensors. Remember that the action of a basis vector e_i on a dual vector e^j is given by $e_i(e^j) = e^j(e_i) = \delta_i^j$. In our case this means that

$$\frac{\partial}{\partial x^i}(dx^j) = dx^j(\frac{\partial}{\partial x^i}) = \delta_i^j$$

The general form of a contravariant tensor is therefore

$$T = \sum_{j_1 \ldots j_p}^{n} T^{j_1 \ldots j_p} \frac{\partial}{\partial x^{j_1}} \otimes \ldots \otimes \frac{\partial}{\partial x^{j_p}}$$

For the transformation law we have

$$\begin{aligned} T &= \sum_{j_1 \ldots j_p}^{n} T'^{j_1 \ldots j_p} \frac{\partial}{\partial x'^{j_1}} \otimes \ldots \otimes \frac{\partial}{\partial x'^{j_p}} \\ &= \sum_{j_1 \ldots j_p, k_1 \ldots k_p}^{n} T'^{j_1 \ldots j_p} \frac{\partial x^{k_1}}{\partial x'^{j_1}} \cdots \frac{\partial x^{k_p}}{\partial x'^{j_p}} \frac{\partial}{\partial x^{k_1}} \otimes \ldots \otimes \frac{\partial}{\partial x^{k_p}} \\ &= \sum_{k_1 \ldots k_p}^{n} T^{k_1 \ldots k_n} \frac{\partial}{\partial x^{k_1}} \otimes \ldots \otimes \frac{\partial}{\partial x^{k_p}} \end{aligned}$$

and hence

$$T^{k_1\ldots k_n} = \sum_{j_1\ldots j_p}^{n} T'^{j_1\ldots j_p}\,\frac{\partial x^{k_1}}{\partial x'^{j_1}}\cdots\frac{\partial x^{k_p}}{\partial x'^{j_p}} \tag{3.131}$$

which is a generalization of the transformation law of vectors Eq.(3.125).
Finally we can consider mixed tensors. A mixed tensor which is covariant of order p and contra variant of order q has the form

$$T = \sum_{j_1\ldots j_p, i_1\ldots i_q}^{n} T_{j_1\ldots j_p}{}^{i_1\ldots i_q}\, dx^{j_1}\otimes\ldots\otimes dx^{j_p}\otimes\frac{\partial}{\partial x^{i_1}}\otimes\ldots\otimes\frac{\partial}{\partial x^{i_q}}$$

The transformation law for this tensor is, as you can check for yourself, given by

$$T_{j_1\ldots j_p}{}^{i_1\ldots i_q} = \sum_{k_1\ldots k_p, l_1\ldots l_q} T_{k_1\ldots k_p}{}^{l_1\ldots l_q}\frac{\partial x'^{k_1}}{\partial x^{j_1}}\cdots\frac{\partial x'^{k_p}}{\partial x^{j_p}}\frac{\partial x^{i_1}}{\partial x'^{l_1}}\cdots\frac{\partial x^{i_q}}{\partial x'^{l_q}}$$

A feature that we will often use is how tensors change under mappings between manifolds. Let us first consider the case of convectors. We saw that if we have a map $\phi : M \to N$ from a manifold M to a manifold M then there is a map $\phi_* : T_pM \to T_pN$ called the push-forward, that maps a tangent vector at M to a tangent vector at N. This can used to define a dual map $\phi^* : T_pN^* \to T_pM^*$ that maps a covector on N to a covector on M. If ω is a covector on N, then we define a new covector $\phi^*\omega$ of M by

$$(\phi^*\omega)(v) = \omega(\phi_* v) \tag{3.132}$$

where v is a vector on M. Since $\phi^*\omega$ is a covector on M we say that ω is pulled back by ϕ from N to M. Remember from Chapter 2 (see Eq.(2.63)) that functions defined on manifolds have the same property. If f is a function on N then $\phi^* f = f\circ\phi$ is a function on M. So functions are also pulled back. In fact, we can derive the following useful formula

$$\phi^* df = d(\phi^* f) \tag{3.133}$$

This formula follows directly from all our definitions of pullbacks and push forwards. We have

$$(\phi^* df)(v) = df(\phi_* v) = (\phi_* v)(f) = v(\phi^* f) = v(f\circ\phi) = d(f\circ\phi)(v) = d(\phi^* f)(v)$$

You can check each step by reviewing all our definitions Eqs.(2.63), (3.19), (3.117) and (3.132). As an example we take as ϕ the embedding map of the torus $i : M \to \mathbb{R}^3$ of Eq.(3.112). Take an arbitrary covector

$$\omega = \omega_1 dx^1 + \omega_2 dx^2 + \omega_3 dx^3$$

on $\mathbb{R}^3$. Then

$$\begin{aligned} i^*\omega &= \omega_1 d(i^* x^1) + \omega_2 d(i^* x^2) + \omega_3 d(i^* x^3) \\ &= \omega_1 d(x^1\circ i) + \omega_2 d(x^2\circ i) + \omega_3 d(x^3\circ i) \end{aligned} \tag{3.134}$$

We have

$$\begin{aligned} d(x^1\circ i) &= d((a+b\cos v)\cos u) = -\sin u(a+b\cos v)du - b\sin v\cos u dv \\ d(x^2\circ i) &= d((a+b\cos v)\sin u) = \cos u(a+b\cos v)du - b\sin v\sin u dv \\ d(x^2\circ i) &= d(b\sin v) = b\cos v dv \end{aligned}$$

In matrix notation

$$i^*\omega = \begin{pmatrix} -\sin u(a+b\cos v) & \cos u(a+b\cos v) & 0 \\ -b\sin v\cos u & b\sin v\sin u & b\cos v \end{pmatrix} \begin{pmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \end{pmatrix}$$

This is nothing but the transpose of the matrix in Eq.(3.114) of the push forward mapping i_*. Let us finally extend the pullback and push forward transformations to general tensors. For a general covariant tensor of order q we can define a pullback similarly as in Eq.(3.132). If $\phi : M \to N$ and T is a covariant tensor on N then ϕ^*T is a covariant tensor on M given by

$$(\phi^*T)(p)(v_1,\ldots,v_q) = T(\phi(p))(\phi_*v_1,\ldots,\phi_*v_q) \tag{3.135}$$

where v_j are vectors on M and ϕ maps point p on M to a point $\phi(p)$ of N. Similarly if T is a contravariant tensor of order q on M then ϕ_*T is a contravariant tensor on N given by

$$(\phi_*T)(\phi(p))(\omega_1,\ldots,\omega_q) = T(p)(\phi^*\omega_1,\ldots,\phi^*\omega_q) \tag{3.136}$$

where ω_j are covectors on N. This defines the push forward of contravariant tensors. We will mostly use the pullback operation of Eq.(3.135). For example, if

$$g = \sum_{i,j}^{n} g_{ij}(x)dx^i \otimes dx^j$$

is a metric tensor in coordinate system $x = (x^1,\ldots,x^n)$ on N then for vectors $v_1, v_2 \in T_pM$ we have

$$\begin{aligned}(\phi^*g)(p)(v_1,v_2) &= \sum_{i,j}^{n} g_{ij}(\phi(p))dx^i \otimes dx^j(\phi_*v_1,\phi_*v_2) = \sum_{i,j}^{n} g_{ij}(\phi(p))dx^i(\phi_*v_1)dx^j(\phi_*v_2) \\ &= \sum_{i,j}^{n} g_{ij}(\phi(p))d(\phi^*x^i)(v_1)d(\phi^*x^j)(v_2) \\ &= \sum_{i,j}^{n} g_{ij}(\phi(p))d(\phi^*x^i)\otimes d(\phi^*x^j)(v_1,v_2)\end{aligned} \tag{3.137}$$

We can, for instance, apply this to the embedding of the torus in $\mathbb{R}^3$. We take the Euclidean metric on $\mathbb{R}^3$ given by

$$g = dx^1 \otimes dx^1 + dx^2 \otimes dx^2 + dx^3 \otimes dx^3 \tag{3.138}$$

Then under the embedding mapping i of Eq.(3.112) we have

$$\begin{aligned} i^*g &= d(x^1\circ i)\otimes d(x^1\circ i) + d(x^2\circ i)\otimes d(x^2\circ i) + d(x^3\circ i)\otimes d(x^3\circ i) \\ &= \sum_{j=1}^{3}\left(\frac{\partial x^j}{\partial u}du + \frac{\partial x^j}{\partial v}dv\right)\otimes\left(\frac{\partial x^j}{\partial u}du + \frac{\partial x^j}{\partial v}dv\right) \\ &= g_{uu}\,du\otimes du + g_{uv}\,(du\otimes dv + dv\otimes du) + g_{vv}\,dv\otimes dv \end{aligned} \tag{3.139}$$

where we defined

$$\begin{aligned} g_{uu} &= \sum_{j=1}^{3} \frac{\partial x^j}{\partial u}\frac{\partial x^j}{\partial u} = \langle \frac{\partial \mathbf{x}}{\partial u}, \frac{\partial \mathbf{x}}{\partial u} \rangle \\ g_{uv} &= \sum_{j=1}^{3} \frac{\partial x^j}{\partial u}\frac{\partial x^j}{\partial v} = \langle \frac{\partial \mathbf{x}}{\partial u}, \frac{\partial \mathbf{x}}{\partial v} \rangle \\ g_{vv} &= \sum_{j=1}^{3} \frac{\partial x^j}{\partial v}\frac{\partial x^j}{\partial v} = \langle \frac{\partial \mathbf{x}}{\partial v}, \frac{\partial \mathbf{x}}{\partial v} \rangle \end{aligned} \tag{3.140}$$

and defined the standard inner product

$$\langle a, b \rangle = a^1 b^1 + a^2 b^2 + a^3 b^3$$

on $\mathbb{R}^3$. From Eqs.(3.115) and (3.116) we see that

$$\begin{aligned} g_{uu} &= \langle \frac{\partial \mathbf{x}}{\partial u}, \frac{\partial \mathbf{x}}{\partial u} \rangle = (a + b\cos v)^2 \\ g_{uv} &= \langle \frac{\partial \mathbf{x}}{\partial u}, \frac{\partial \mathbf{x}}{\partial v} \rangle = 0 \\ g_{vv} &= \langle \frac{\partial \mathbf{x}}{\partial v}, \frac{\partial \mathbf{x}}{\partial v} \rangle = b^2 \end{aligned} \tag{3.141}$$

We therefore see that the usual Euclidean inner product restricted to the torus attains the form

$$i^* g = (a + b\cos v)^2 du \otimes du + b^2 dv \otimes dv \tag{3.142}$$

A similar procedure can be carried out for calculating the components of the metric tensor after a coordinate transformation. After all, a coordinate transformation is just a special case of a mapping $\phi : M \to N$ where $M = N$. We have

$$g'_{kl} = \sum_{ij}^{n} g_{ij} \frac{\partial x^i}{\partial x'^k}\frac{\partial x^j}{\partial x'^l} = \langle \frac{\partial \mathbf{x}}{\partial x'^k}, \frac{\partial \mathbf{x}}{\partial x'^l} \rangle$$

where we defined

$$\langle a, b \rangle = \sum_{i,j}^{n} g_{ij} a^i b^j$$

If we take (x^1, x^2, x^3) to be the Cartesian coordinates on $\mathbb{R}^3$ with the usual Euclidean metric $g_{ij} = \delta_{ij}$ and take $(x'^1, x'^2, x'^3) = (r, \phi, \theta)$ to be the spherical coordinates

$$\begin{aligned} x^1 &= r\cos\phi\sin\theta \\ x^2 &= r\sin\phi\sin\theta \\ x^3 &= r\cos\theta \end{aligned}$$

then we have

$$\frac{\partial \mathbf{x}}{\partial \phi} = \begin{pmatrix} -r\sin\phi\sin\theta \\ r\cos\phi\sin\theta \\ 0 \end{pmatrix} \quad , \quad \frac{\partial \mathbf{x}}{\partial \theta} = \begin{pmatrix} r\cos\phi\cos\theta \\ r\sin\phi\cos\theta \\ -r\sin\theta \end{pmatrix} \quad , \quad \frac{\partial \mathbf{x}}{\partial r} = \begin{pmatrix} \cos\phi\sin\theta \\ \sin\phi\sin\theta \\ \cos\theta \end{pmatrix}$$

Then

$$
\begin{aligned}
g_{\phi\phi} &= \langle \frac{\partial \mathbf{x}}{\partial \phi}, \frac{\partial \mathbf{x}}{\partial \phi} \rangle = r^2 \sin^2 \theta \\
g_{\theta\theta} &= \langle \frac{\partial \mathbf{x}}{\partial \theta}, \frac{\partial \mathbf{x}}{\partial \theta} \rangle = r^2 \\
g_{rr} &= \langle \frac{\partial \mathbf{x}}{\partial r}, \frac{\partial \mathbf{x}}{\partial r} \rangle = 1
\end{aligned}
\tag{3.143}
$$

and $g_{r\phi} = g_{r\theta} = g_{\phi\theta} = 0$. We therefore find that the metric tensor in spherical coordinates attains the form

$$g = dr \otimes dr + r^2 d\theta \otimes d\theta + r^2 \sin^2 \theta d\phi \otimes d\phi \tag{3.144}$$

In a similar manner you can check that the metric of the unit sphere ($r = 1$) inherits the metric

$$i^* g = d\theta \otimes d\theta + \sin^2 \theta d\phi \otimes d\phi$$

by embedding in $\mathbb{R}^3$ with the usual Euclidean inner product.

Chapter 4

Motion of particles

We use the Lagrangian principle to write Newton's law of motion in general coordinates and deduce its transformation law to different coordinates. We show that particles which move freely restricted to a two-dimensional surface move at constant speed along geodesic curves. We further give a condition that tells us whether a given metric can be reduced to Euclidean form, which leads us to the introduction of the Riemann tensor.

4.1 Lagrangian equations

The goal of this section is to write Newton's law

$$m\frac{d^2x^i}{dt^2} = -\frac{\partial V}{\partial x^i} \tag{4.1}$$

in such a way that it is valid in any spatial coordinate system. The key step in doing this is by using the Lagrangian formulation of classical mechanics. We define the Lagrangian to be

$$L(x, \dot{x}) = T - V = \sum_{i=1}^{3} \frac{1}{2}m(\dot{x}^i)^2 - V(x^1, x^2, x^3) \tag{4.2}$$

which is just the difference between the kinetic energy T and the potential energy T. We will further use the notation $\dot{x}^i = dx^i/dt$ and $\ddot{x}^i = d^2x^i/dt^2$ to compactify our notation. We then define the action functional to be

$$S[x] = \int_{t_0}^{t_1} dt\, L(x, \dot{x}) \tag{4.3}$$

where the square brackets denote a functional dependence on the path. Eq.(4.1) then follows as the stationary point of the action, i.e. as the equation for the trajectory x that is a local minimum, maximum, or saddle point of S subject to the constraint that the endpoints $x(t_0)$ and $x(t_1)$ of the path $x(t)$ are fixed.

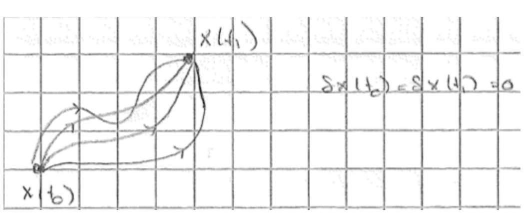

Figure 4.1: Different paths subject to the condition $\delta x(t_0) = \delta x(t_1) = 0$.

This means the following. If $x(t)$ is a stationary path for $S[x]$ then any other path $x(t)+\epsilon\delta x(t)$ with $\delta x(t_0)=\delta x(t_1)=0$ will to first order in ϵ not change the value of $S[x]$. This implies that

$$\frac{\partial}{\partial\epsilon}S[x+\epsilon\delta x]=0$$

or equivalently

$$0=\lim_{\epsilon\to 0}\frac{S[x+\epsilon\delta x]-S[x]}{\epsilon} \tag{4.4}$$

Let us apply this condition to Eq.(4.2). We have

$$\begin{aligned}
S[x+\epsilon\delta x]-S[x] &= \int_{t_0}^{t_1} dt(L(x+\epsilon\delta x,\dot{x}+\epsilon\delta\dot{x})-L(x,\dot{x})) \\
&= \epsilon\sum_{j=1}^{3}\int_{t_0}^{t_1} dt\left(\frac{\partial L}{\partial x^i}\delta x^i(t)+\frac{\partial L}{\partial \dot{x}^i}\delta\dot{x}^i(t)\right)+O(\epsilon^2) \\
&= \epsilon\sum_{j=1}^{3}\int_{t_0}^{t_1} dt\delta x^i(t)\left[\frac{\partial L}{\partial x^i}-\frac{\partial}{\partial t}\frac{\partial L}{\partial \dot{x}^i}\right]+O(\epsilon^2)
\end{aligned}$$

where we used the boundary condition $\delta x(t_0)=\delta x(t_1)=0$ when we performed the partial integration. Then we see that we must have the condition

$$0=\int_{t_0}^{t_1} dt\delta x^i(t)\left[\frac{\partial L}{\partial x^i}-\frac{\partial}{\partial t}\frac{\partial L}{\partial \dot{x}^i}\right] \tag{4.5}$$

in order to satisfy Eq.(4.5). Since Eq.(4.5) must be true for any variation $\delta x^i(t)$ the stationary path must satisfy

$$0=\frac{\partial L}{\partial x^i}-\frac{\partial}{\partial t}\frac{\partial L}{\partial \dot{x}^i} \tag{4.6}$$

If we take the Lagrangian L to be equal to Eq.(4.2) then we find

$$\frac{\partial L}{\partial x^i}=-\frac{\partial V}{\partial x^i}\quad , \quad \frac{\partial L}{\partial \dot{x}^i}=m\dot{x}^i$$

and we find that Eq.(4.6) yields the equation of motion of Eq.(4.1). The main reason that we went through all this trouble in deriving Newton's law once again is that the Lagrangian principle can be applied in any coordinate system. If we go from coordinates x^i to new coordinates y^i then from differentiation of $x^i(y^1,\ldots,y^n)$ we have

$$\dot{x}^i=\sum_{k=1}^{n}\frac{\partial x^i}{\partial y^k}\dot{y}^k$$

and consequently

$$\sum_{i=1}^{n}(\dot{x}^i)^2=\sum_{i,k,l=1}^{n}\frac{\partial x^i}{\partial y^k}\frac{\partial x^i}{\partial y^l}\dot{y}^k\dot{y}^l=\sum_{k,l=1}^{n}g_{kl}\dot{y}^k\dot{y}^l$$

where we defined

$$g_{kl}=\sum_{i,j=1}^{n}\delta_{ij}\frac{\partial x^i}{\partial y^k}\frac{\partial x^j}{\partial y^l} \tag{4.7}$$

which represents a transformation of the Euclidean metric to general coordinates. As a consequence Eq.(4.2) attains the form

$$L(y,\dot y) = \sum_{k,l}^{3} \frac{1}{2} m\, g_{kl}\, \dot y^k \dot y^l - V(y_1,y_2,y_3) \tag{4.8}$$

where, with some abuse of notation, we wrote $V(y_1,y_2,y_3) = V(x^1(y),x^2(y),x^3(y))$. The action van now be written as

$$S[y] = \int_{t_0}^{t_1} dt\, L(y,\dot y) \tag{4.9}$$

Obviously, since Eq.(4.9) is identical to Eq.(4.3) the stationary path will be identical. It will just be expressed in different coordinates. For instance, in spherical coordinates we have

$$L = \frac{1}{2} m\,(\dot r^2 + r^2\,\dot\theta^2 + r^2 \sin^2\theta\,\dot\phi^2) - V(r,\phi,\theta) \tag{4.10}$$

This Lagrangian is very useful for problems of spherical symmetry, such as planetary motion around the Sun. Let us see how the equations of motion in general coordinates look like. We write

$$S[x] = \int_{t_0}^{t_1} dt \left(\sum_{i,j}^{n} \frac{1}{2} m\, g_{ij}\, \dot x^i \dot x^j - V(x) \right) \tag{4.11}$$

where x^i now represent general coordinates. We have

$$\begin{aligned} \frac{\partial L}{\partial x^k} &= -\frac{\partial V}{\partial x^k} + \frac{1}{2} m \sum_{i,j=1}^{3} \frac{\partial g_{ij}}{\partial x^k} \dot x^i \dot x^j \\ \frac{\partial L}{\partial \dot x^k} &= \frac{1}{2} m \sum_{i,j=1}^{n} g_{ij} (\delta_{ik} \dot x^j + \dot x^i \delta_{jk}) = m \sum_{j=1}^{n} g_{kj} \dot x^j \end{aligned}$$

Eq.(4.6) then gives

$$0 = \frac{\partial}{\partial t} \frac{\partial L}{\partial \dot x^i} - \frac{\partial L}{\partial x^i} = m \frac{\partial}{\partial t} \left(\sum_{j=1}^{n} g_{kj} \dot x^j \right) + \frac{\partial V}{\partial x^k} - \frac{1}{2} m \sum_{i,j=1}^{3} \frac{\partial g_{ij}}{\partial x^k} \dot x^i \dot x^j$$

This can be rewritten as

$$m \sum_{i,j=1}^{n} \frac{\partial g_{kj}}{\partial x^i} \dot x^i \dot x^j + m \sum_{j=1}^{n} g_{kj} \ddot x^j - \frac{1}{2} m \sum_{i,j=1}^{3} \frac{\partial g_{ij}}{\partial x^k} \dot x^i \dot x^j = -\frac{\partial V}{\partial x^k}$$

and with the notation

$$[ij,k] = \frac{1}{2} \left(\frac{\partial g_{ik}}{\partial x^j} + \frac{\partial g_{jk}}{\partial x^i} - \frac{\partial g_{ij}}{\partial x^k} \right) \tag{4.12}$$

as

$$m \sum_{j=1}^{n} g_{kj} \ddot x^j + m \sum_{i,j=1}^{n} [ij,k]\, \dot x^i \dot x^j = -\frac{\partial V}{\partial x^k} \tag{4.13}$$

the symbols $[ij,k]$ are called Christoffel symbols of the first kind. Eq.(4.13) can be further rewritten by multiplication with g^{lk} and carrying out a summation over k. this yields

$$m \left(\ddot x^l + \sum_{i,j=1}^{n} \Gamma^l_{ij}\, \dot x^i \dot x^j \right) = -\sum_{k=1}^{n} g^{lk} \frac{\partial V}{\partial x^k} \tag{4.14}$$

where we defined

$$\Gamma^l_{ij} = \sum_{k=1}^{n} g^{lk}[ij,k] \tag{4.15}$$

where Γ^l_{ij} are called Christoffel symbols of the second kind. The right hand side of Eq.(4.13) transforms like a covector (see Eq.(3.127)) and hence so must the left hand side. In Eq.(4.14) we transformed the right hand side to vector by raising an index, and therefore the left hand side must also transform like a vector. This fact leads to a specific transformation law for the Christoffel symbols Γ^k_{ij}. If we change coordinates to y^α with $\alpha = 1, \ldots, n$ we have

$$\begin{aligned}
\dot{x}^i &= \sum_{\alpha}^{n} \frac{\partial x^i}{\partial y^\alpha} \dot{y}^\alpha \\
\ddot{x}^i &= \sum_{\alpha}^{n} \frac{\partial x^i}{\partial y^\alpha} \ddot{y}^\alpha + \sum_{\alpha,\beta}^{n} \frac{\partial^2 x^i}{\partial y^\alpha \partial y^\beta} \dot{y}^\alpha \dot{y}^\beta
\end{aligned}$$

Inserting these expressions into Eq.(4.14) then yields

$$\begin{aligned}
m \sum_{\alpha}^{n} \frac{\partial x^l}{\partial y^\alpha} \ddot{y}^\alpha &+ m \sum_{\alpha,\beta,i,j}^{n} \dot{y}^\alpha \dot{y}^\beta \left(\Gamma^l_{ij} \frac{\partial x^i}{\partial y^\alpha} \frac{\partial x^j}{\partial y^\beta} + \frac{\partial^2 x^l}{\partial y^\alpha \partial y^\beta} \right) \\
&= - \sum_{k,\alpha,\beta,\gamma} g'^{\alpha\beta} \frac{\partial x^l}{\partial y^\alpha} \frac{\partial x^k}{\partial y^\beta} \frac{\partial y^\gamma}{\partial x^k} \frac{\partial V}{\partial y^\gamma} \\
&= - \sum_{\alpha,\beta,\gamma} g'^{\alpha\beta} \frac{\partial x^l}{\partial y^\alpha} \delta^\gamma_\beta \frac{\partial V}{\partial y^\gamma} = - \sum_{\alpha,\beta} \frac{\partial x^l}{\partial y^\alpha} g'^{\alpha\beta} \frac{\partial V}{\partial y^\beta}
\end{aligned} \tag{4.16}$$

where we used

$$\sum_{k}^{n} \frac{\partial x^k}{\partial y^\beta} \frac{\partial y^\gamma}{\partial x^k} = \frac{\partial y^\gamma}{\partial y^\beta} = \delta^\gamma_\beta$$

and where $g'^{\alpha\beta}$ is the inverse metric tensor in coordinates system y. Multiplying Eq.(4.16) by $\partial y^\gamma / \partial x^l$ and summing over l then yields

$$m\ddot{y}^\gamma + \sum_{\alpha,\beta,l,i,j}^{n} \dot{y}^\alpha \dot{y}^\beta \left(\Gamma^l_{ij} \frac{\partial x^i}{\partial y^\alpha} \frac{\partial x^j}{\partial y^\beta} \frac{\partial y^\gamma}{\partial x^l} + \frac{\partial^2 x^l}{\partial y^\alpha \partial y^\beta} \frac{\partial y^\gamma}{\partial x^l} \right) = - \sum_{\beta}^{n} g'^{\gamma\beta} \frac{\partial V}{\partial y^\beta} \tag{4.17}$$

This equation can be rewritten as

$$m \left(\ddot{y}^\gamma + \sum_{\alpha,\beta,}^{n} \Gamma'^{\gamma}_{\alpha\beta} \dot{y}^\alpha \dot{y}^\beta \right) = - \sum_{\beta}^{n} g'^{\gamma\beta} \frac{\partial V}{\partial y^\beta} \tag{4.18}$$

provided that

$$\Gamma'^{\gamma}_{\alpha\beta} = \sum_{i,j,l}^{n} \left(\Gamma^l_{ij} \frac{\partial x^i}{\partial y^\alpha} \frac{\partial x^j}{\partial y^\beta} \frac{\partial y^\gamma}{\partial x^l} + \frac{\partial^2 x^l}{\partial y^\alpha \partial y^\beta} \frac{\partial y^\gamma}{\partial x^l} \right) \tag{4.19}$$

Since we know that Eq.(4.14) is valid in any coordinate system the Christoffel symbols must in fact transform according to Eq.(4.19). It is not difficult to check this explicitly. We start by calculating

$$\frac{\partial g'_{\alpha\beta}}{\partial y^\gamma} = \frac{\partial}{\partial y^\gamma}\left(\sum_{i,j}^n g_{ij}\frac{\partial x^i}{\partial y^\alpha}\frac{\partial x^j}{\partial y^\beta}\right) = \sum_{i,j,k}^n \frac{\partial g_{ij}}{\partial x^k}\frac{\partial x^i}{\partial y^\alpha}\frac{\partial x^j}{\partial y^\beta}\frac{\partial x^k}{\partial y^\gamma}$$
$$+\sum_{i,j}^n g_{ij}\left(\frac{\partial x^i}{\partial y^\alpha}\frac{\partial^2 x^j}{\partial y^\beta \partial y^\gamma} + \frac{\partial x^j}{\partial y^\beta}\frac{\partial^2 x^i}{\partial y^\alpha \partial y^\gamma}\right)$$

If we use this equation then from Eq.(4.12) it follows that

$$\begin{aligned}[\alpha\beta,\gamma]' =&\frac{1}{2}\left(\frac{\partial g'_{\alpha\gamma}}{\partial y^\beta} + \frac{\partial g'_{\beta\gamma}}{\partial y^\alpha} - \frac{\partial g'_{\alpha\beta}}{\partial y^\gamma}\right)\\ =&\frac{1}{2}\sum_{i,j,k}^n \frac{\partial g_{ij}}{\partial x^k}\left(\frac{\partial x^i}{\partial y^\alpha}\frac{\partial x^j}{\partial y^\gamma}\frac{\partial x^k}{\partial y^\beta} + \frac{\partial x^i}{\partial y^\beta}\frac{\partial x^j}{\partial y^\gamma}\frac{\partial x^k}{\partial y^\alpha} - \frac{\partial x^i}{\partial y^\alpha}\frac{\partial x^j}{\partial y^\beta}\frac{\partial x^k}{\partial y^\gamma}\right)\\ &+\frac{1}{2}\sum_{i,j}^n g_{ij}\left(\frac{\partial x^i}{\partial y^\alpha}\frac{\partial^2 x^j}{\partial y^\gamma \partial y^\beta} + \frac{\partial x^j}{\partial y^\gamma}\frac{\partial^2 x^i}{\partial y^\alpha \partial y^\beta} + \frac{\partial x^i}{\partial y^\beta}\frac{\partial^2 x^j}{\partial y^\gamma \partial y^\alpha} + \frac{\partial x^j}{\partial y^\gamma}\frac{\partial^2 x^i}{\partial y^\beta \partial y^\alpha}\right.\\ &\left.-\frac{\partial x^i}{\partial y^\alpha}\frac{\partial^2 x^j}{\partial y^\beta \partial y^\gamma} - \frac{\partial x^j}{\partial y^\beta}\frac{\partial^2 x^i}{\partial y^\alpha \partial y^\gamma}\right)\end{aligned}$$

After some relabeling of indices this is then rewritten as

$$\begin{aligned}[\alpha\beta,\gamma]' =&\frac{1}{2}\sum_{i,j,k}^n \left(\frac{\partial g_{ik}}{\partial x^j} + \frac{\partial g_{jk}}{\partial x^i} - \frac{\partial g_{ij}}{\partial x^k}\right)\frac{\partial x^i}{\partial y^\alpha}\frac{\partial x^j}{\partial y^\beta}\frac{\partial x^k}{\partial y^\gamma} + \sum_{i,j}^n g_{ij}\frac{\partial x^i}{\partial y^\gamma}\frac{\partial^2 x^j}{\partial y^\alpha \partial y^\beta}\\ =&\frac{1}{2}\sum_{i,j,k}^n [ij,k]\frac{\partial x^i}{\partial y^\alpha}\frac{\partial x^j}{\partial y^\beta}\frac{\partial x^k}{\partial y^\gamma} + \sum_{i,j}^n g_{ij}\frac{\partial x^i}{\partial y^\gamma}\frac{\partial^2 x^j}{\partial y^\alpha \partial y^\beta}\end{aligned} \tag{4.20}$$

Using Eq.(4.20) we then have

$$\begin{aligned}\Gamma'^{\gamma}_{\alpha\beta} =&\sum_{\delta}^n g'^{\gamma\delta}[\alpha\beta,\delta]' = \sum_{p,q,\delta}^n g^{pq}\frac{\partial y^\gamma}{\partial x^p}\frac{\partial y^\delta}{\partial x^q}[\alpha\beta,\delta]'\\ =&\sum_{i,j,k,p,q,\delta}^n g^{pq}[ij,k]\frac{\partial y^\gamma}{\partial x^p}\underbrace{\frac{\partial y^\delta}{\partial x^q}\frac{\partial x^k}{\partial y^\delta}}_{\delta^k_q}\frac{\partial x^i}{\partial y^\alpha}\frac{\partial x^j}{\partial y^\beta} + \sum_{i,j,p,q,\delta}^n g^{pq}g_{ij}\frac{\partial y^\gamma}{\partial x^p}\underbrace{\frac{\partial y^\delta}{\partial x^q}\frac{\partial x^i}{\partial y^\delta}}_{\delta^i_q}\frac{\partial^2 x^j}{\partial y^\alpha \partial y^\beta}\\ =&\sum_{i,j,p,q}^n g^{pq}[ij,q]\frac{\partial x^i}{\partial y^\alpha}\frac{\partial x^j}{\partial y^\beta}\frac{\partial y^\gamma}{\partial x^p} + \sum_{p,q,j}\underbrace{g^{pq}g_{qj}}_{\delta^p_j}\frac{\partial y^\gamma}{\partial x^p}\frac{\partial^2 x^j}{\partial y^\alpha \partial y^\beta}\\ =&\sum_{i,j,p}^n \Gamma^p_{ij}\frac{\partial x^i}{\partial y^\alpha}\frac{\partial x^j}{\partial y^\beta}\frac{\partial y^\gamma}{\partial x^p} + \sum_{j}^n \frac{\partial^2 x^j}{\partial y^\alpha \partial y^\beta}\frac{\partial y^\gamma}{\partial x^j}\end{aligned} \tag{4.21}$$

which is exactly the transformation law of Eq.(4.19).
The form of the Langrangian (4.8) is unavoidable for the description of particle motion constrained to surfaces since in that case no metric of Euclidean form is available. For a general surface parametrized by $(x^1(u,v), x^2(u,v), x^3(u,v))$ we have

$$T = \frac{1}{2}m\sum_{i=1}^3 (\dot{x}^i)^2 = \frac{1}{2}m\sum_{i=1}^3 (\frac{\partial x^i}{\partial u}\dot{u} + \frac{\partial x^i}{\partial v}\dot{v})^2 = \frac{1}{2}m(g_{uu}\dot{u}^2 + 2g_{uv}\dot{u}\dot{v} + g_{vv}\dot{v}^2) \tag{4.22}$$

where g_{uu}, g_{vv} and g_{uv} are given by the expressions (3.140). In particular, for particles on a torus we have from Eq.(3.141) that

$$L = \frac{1}{2}m((a + b\cos v)^2\dot{u}^2 + b^2\dot{v}^2) - V(u, v) \tag{4.23}$$

For the equations of motion we have

$$\begin{aligned}
0 &= \frac{\partial L}{\partial u} - \frac{\partial}{\partial t}\frac{\partial L}{\partial \dot{u}} = -\frac{\partial V}{\partial u} - \frac{\partial}{\partial t}(m(a + b\cos v)^2\dot{u}) \\
0 &= \frac{\partial L}{\partial v} - \frac{\partial}{\partial t}\frac{\partial L}{\partial \dot{v}} = -\frac{\partial V}{\partial v} - m\,\dot{u}^2(a + b\cos v)b\sin v - mb^2\ddot{v}
\end{aligned}$$

which can be rewritten as

$$\begin{aligned}
m(a + b\cos v)^2\,\ddot{u} - 2mb\sin v(a + b\cos v)\,\dot{u}\,\dot{v} &= -\frac{\partial V}{\partial u} \\
mb^2\,\ddot{v} + mb\sin v(a + b\cos v)\,\dot{u}^2 &= -\frac{\partial V}{\partial v}
\end{aligned}$$

These equations are equivalent to Eq.(4.13). Multiplying with the inverse metric (which is easy since the metric is diagonal) we have

$$m\left[\ddot{u} - \frac{2b\sin v}{a + b\cos v}\,\dot{u}\,\dot{v}\right] = -\frac{1}{(a + b\cos v)^2}\frac{\partial V}{\partial u} \tag{4.24}$$

$$m\left[\ddot{v} + \frac{1}{b}\sin v\,(a + b\cos v)\,\dot{u}^2\right] = -\frac{1}{b^2}\frac{\partial V}{\partial v} \tag{4.25}$$

These equations are equivalent to Eq.(4.14). From these equations we can directly read off the Christoffel symbols Γ^k_{ij}. We have

$$\begin{aligned}
\Gamma^u_{uv} &= \Gamma^u_{vu} = -\frac{b\sin v}{a + b\cos v} \\
\Gamma^v_{uu} &= \frac{1}{b}\sin v\,(a + b\cos v)
\end{aligned}$$

and all other Christoffel symbols are zero. We can check this using Eqs.(4.12) and (4.15). We have

$$[uv, u] = \frac{1}{2}\left(\frac{\partial g_{uu}}{\partial v} + \frac{\partial g_{vu}}{\partial u} - \frac{\partial g_{uv}}{\partial u}\right) = \frac{1}{2}\frac{\partial}{\partial v}(a + b\cos v)^2 = -b\sin v\,(a + b\cos v) \tag{4.26}$$

$$[uu, v] = \frac{1}{2}\left(\frac{\partial g_{uv}}{\partial u} + \frac{\partial g_{uv}}{\partial u} - \frac{\partial g_{uu}}{\partial v}\right) = -\frac{1}{2}\frac{\partial}{\partial v}(a + b\cos v)^2 = b\sin v\,(a + b\cos v) \tag{4.27}$$

and therefore

$$\begin{aligned}
\Gamma^u_{uv} &= g^{uu}[uv, u] = -\frac{b\sin v(a + b\cos v)}{(a + b\cos v)^2} = -\frac{b\sin v}{(a + b\cos v)} \\
\Gamma^v_{uu} &= g^{vv}[uu, v] = \frac{1}{b}\sin v(a + b\cos v)
\end{aligned}$$

which gives the same result as before.

4.2 Energy conservation and geodesic motion

Let us further check another well known feature of Newton's equations, namely the conservation of energy (provided that the potential V does not explicitly depend on time). Let us take the time-derivative of the kinetic energy

$$\frac{\partial T}{\partial t} = \frac{\partial}{\partial t}\left(\frac{1}{2}m\sum_{i,j}^{n} g_{ij}\dot{x}^i\dot{x}^j\right) = \frac{1}{2}m\sum_{i,j,k}^{n}\frac{\partial g_{ij}}{\partial x^k}\dot{x}^i\dot{x}^j\dot{x}^k + \frac{1}{2}m\sum_{i,j}^{n} g_{ij}(\ddot{x}^i\dot{x}^j + \dot{x}^i\ddot{x}^j) \tag{4.28}$$

From Eq.(4.12) we see that

$$\frac{\partial g_{ij}}{\partial x^k} = [ik,j] + [jk,i]$$

If we insert this expression into Eq.(4.28) we obtain

$$\begin{aligned}\frac{\partial T}{\partial t} &= \frac{1}{2}m\sum_{j}^{n}\dot{x}^j\left[\sum_{i}^{n} g_{ij}\ddot{x}^i + \sum_{i,k}^{n}[ik,j]\dot{x}^i\dot{x}^k\right] + \frac{1}{2}m\sum_{i}^{n}\dot{x}^i\left[\sum_{j}^{n} g_{ij}\ddot{x}^j + \sum_{j,k}^{n}[jk,i]\dot{x}^j\dot{x}^k\right] \\ &= -\frac{1}{2}\sum_{j}^{n}\dot{x}^j\frac{\partial V}{\partial x^j} - \frac{1}{2}\sum_{i}^{n}\dot{x}^i\frac{\partial V}{\partial x^i} = -\frac{\partial V}{\partial t}\end{aligned} \tag{4.29}$$

where we used the equations of motion (4.13). We therefore find that

$$\frac{\partial}{\partial t}(T+V) = 0 \tag{4.30}$$

and therefore the total energy $E = T + V$ is conserved.
A special case arises when we take $V = 0$ and therefore consider free particle motion restricted to the surface. Eq.(4.22) and (4.30) then tell us that

$$E = \frac{1}{2}m|\mathbf{v}|^2 = \frac{1}{2}m\sum_{j=1}^{3}(\dot{x}^j)^2 = \sum_{i,j=1}^{2} g_{ij}\dot{y}^i\dot{y}^j \tag{4.31}$$

is a constant where $x^j(y^1, y^2)$ are the coordinates on $\mathbb{R}^3$ and (y^1, y^2) are the surface coordinates (see Eq.(4.22)). As one expects the speed $|\mathbf{v}|$ of a freely moving particle is constant. Its path is, however, not a straight line in three-dimensional space but a curve on the surface of the specific manifold that we are considering. What is the shape of this curve? We will show that this curve is a *geodesic*, which is the shortest path connecting two points on the surface. Therefore, freely moving particles move along geodesics. To see this we first have to calculate the length of a curve.

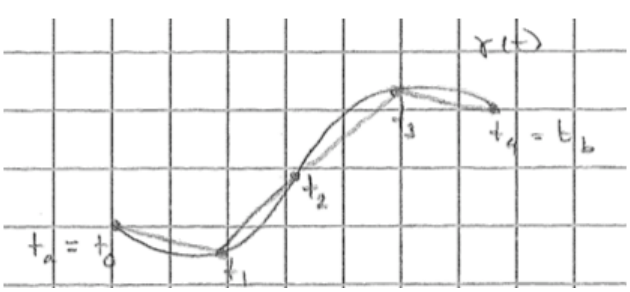

Figure 4.2: Segmentation of a path.

Suppose a particle moves along a curve $\gamma(t)$ in a three-dimensional Cartesian coordinate system. To measure the length of the curve between two points $\gamma(t_a)$ and $\gamma(t_b)$ we consider the position of the particle at different times $t_j = t_a + j\Delta t$, where $\Delta t = (t_b - t_a)/n$ and approximate the path with linear segments as displayed in Fig. 4.2. The length of the segmented path is then given by

$$l_{ab} = \sum_{j=0}^{n-1} |\mathbf{x}(t_j + \Delta t) - \mathbf{x}(t_j)|$$

where

$$|\mathbf{x}| = \sqrt{(x^1)^2 + (x^2)^2 + (x^3)^2}$$

In the limit $n \to \infty$ this sum becomes an integral

$$l_{ab} = \lim_{n\to\infty} \sum_{j=0}^{n-1} \frac{|\mathbf{x}(t_j + \Delta t) - \mathbf{x}(t_j)|}{\Delta t} \Delta t = \int_{t_a}^{t_b} dt\, |\frac{d\mathbf{x}}{dt}| = \int_{t_a}^{t_b} dt \sqrt{(\dot{x}^1)^2 + (\dot{x}^2)^2 + (\dot{x}^3)^2}$$

In case the motion is restricted to a surface we can use Eq. (4.22) to write

$$l_{ab} = \int_{t_a}^{t_b} dt \left(\sum_{i,j=1}^{2} g_{ij} \dot{y}^i \dot{y}^j \right)^{1/2} \tag{4.32}$$

where (y^1, y^2) are surface coordinates (our derivation was is, of course, also valid for higher dimensional surfaces in $\mathbb{R}^n$ as you can readily check yourself). The quantity

$$s(t) = \int_{t_a}^{t} d\bar{t} \left(\sum_{i,j=1}^{2} g_{ij} \dot{y}^i \dot{y}^j \right)^{1/2} = \int_{t_a}^{t} d\bar{t}\, |\frac{d\mathbf{x}}{d\bar{t}}| \tag{4.33}$$

is called the *arc length* and represents the distance travelled between times t_a and t (in this equation we denoted $\dot{y}^i = dy^i/d\bar{t}$). Since there is a $1-1$-correspondence between the time t and the distance $s(t)$ travelled, we can also use s to parametrize the path γ rather than t. In that case we have

$$s_b = s(t_b) = \int_{t_a}^{t_b} dt\, |\frac{d\mathbf{x}}{dt}| = \int_{t_a}^{t_b} dt\, |\frac{d\mathbf{x}}{ds}| \frac{ds}{dt} = \int_{s_a}^{s_b} ds\, |\frac{d\mathbf{x}}{ds}|$$

Differentiating both sides of this equation with respect to s_b gives

$$1 = |\frac{d\mathbf{x}}{ds}| \tag{4.34}$$

So for a curve parametrized by arc length the length of the tangent vector is always equal to one. The relation between t and s is particularly simple for a freely moving particle since in that case $|\mathbf{v}| = |d\mathbf{x}/dt|$ is constant (see Eq.(4.31)) and then Eq.(4.33) gives

$$s(t) = \int_{t_a}^{t} d\bar{t} |\mathbf{v}| = |\mathbf{v}|(t - t_a)$$

which, of course, was to be expected for a particle with constant speed. Let us now derive the equation for the curve with the shortest length between two points. We then need to find the stationary point of the functional

$$l[y] = \int_{t_a}^{t_b} dt \sqrt{\mathcal{L}(y, \dot{y})} \tag{4.35}$$

where

$$\mathcal{L}(y,\dot{y}) = \sum_{i,j}^{n} g_{ij}\,\dot{y}^i\,\dot{y}^j \tag{4.36}$$

This is just a special form of a Lagrangian $L = \sqrt{\mathcal{L}}$ where we again fix the beginning and endpoints of the path. Therefore the variational equations are as in Eq.(4.6) and we find

$$\begin{aligned} 0 &= \frac{\partial\sqrt{\mathcal{L}}}{\partial y^i} - \frac{\partial}{\partial t}\frac{\partial\sqrt{\mathcal{L}}}{\partial \dot{y}^i} = \frac{1}{2\sqrt{\mathcal{L}}}\frac{\partial \mathcal{L}}{\partial y^i} - \frac{\partial}{\partial t}\left(\frac{1}{2\sqrt{\mathcal{L}}}\frac{\partial \mathcal{L}}{\partial \dot{y}^i}\right) \\ &= \frac{1}{2\sqrt{\mathcal{L}}}\left(\frac{\partial \mathcal{L}}{\partial y^i} - \frac{\partial}{\partial t}\frac{\partial \mathcal{L}}{\partial \dot{y}^i}\right) + \frac{1}{4\mathcal{L}^{3/2}}\frac{\partial \mathcal{L}}{\partial t}\frac{\partial \mathcal{L}}{\partial \dot{y}^i} \end{aligned}$$

which can be rewritten as

$$0 = \frac{\partial \mathcal{L}}{\partial y^i} - \frac{\partial}{\partial t}\frac{\partial \mathcal{L}}{\partial \dot{y}^i} + \frac{1}{2\mathcal{L}}\frac{\partial \mathcal{L}}{\partial t}\frac{\partial \mathcal{L}}{\partial \dot{y}^i} \tag{4.37}$$

For a freely moving particle $\mathcal{L}$ represents the kinetic energy up to a multiplicative factor and hence

$$0 = \frac{\partial \mathcal{L}}{\partial y^i} - \frac{\partial}{\partial t}\frac{\partial \mathcal{L}}{\partial \dot{y}^i} \quad , \quad 0 = \frac{\partial \mathcal{L}}{\partial t} \tag{4.38}$$

as a consequence of the equations of motion and the conservation of the energy Eq.(4.31). So the stationary paths of the action of a free particle are also the paths that make the length functional (4.35) stationary and therefore represent geodesic paths. The opposite is also true. We first observe that

$$\frac{1}{2\mathcal{L}}\frac{\partial \mathcal{L}}{\partial t} = \frac{1}{\sqrt{\mathcal{L}}}\frac{\partial\sqrt{\mathcal{L}}}{\partial t} = \frac{1}{|\frac{d\mathbf{x}}{dt}|}\frac{\partial}{\partial t}|\frac{d\mathbf{x}}{dt}| = \frac{1}{\frac{ds}{dt}}\frac{\partial}{\partial t}\left(\frac{ds}{dt}\right) \tag{4.39}$$

where in the last step we used Eq.(4.33). Let $y^i(t)$ be a solution to Eq.(4.37). Then we can reparametrize to $y^i(s)$ where s is the arc length. The curve will still be a stationary point but now the last term in Eq.(4.39) vanishes since for $t = s$ we have $ds/dt = 1$. Consequently the last term in Eq.(4.37) vanishes. But this means that the Eqs.(4.38) are satisfied and that $y^i(s)$ is a stationary point of the action for a free particle. This means that any geodesic is the path of a freely moving particle. The general conclusion of all this analysis is therefore that freely moving particles move along geodesics with constant speed.

Let us discuss this more specifically for the case of a particle on the torus. If we put $V = 0$ it follows from Eqs.(4.24) and (4.25) that

$$\ddot{u} = \frac{2b\sin v}{a + b\cos v}\,\dot{u}\,\dot{v} \tag{4.40}$$

$$\ddot{v} = -\frac{1}{b}\sin v\,(a + b\cos v)\,\dot{u}^2 \tag{4.41}$$

Furthermore from Eq.(4.23) and energy conservation it follows that

$$E = \frac{1}{2}m((a + b\cos v)^2\dot{u}^2 + b^2\dot{v}^2) \tag{4.42}$$

These equations can be simplified to obtain first order differential equations. From Eq.(4.40) we see that we can integrate to

$$\dot{u} = \frac{K}{(a + b\cos v)^2} \tag{4.43}$$

where K is a constant. You can check that the time-derivative of this expression gives back Eq.(4.40). From this equation we can see that that the coordinate velocity $\dot{u}$ in the u-direction is smaller when the particle is on the outside of the torus ($v = 0$) than when it is on the inside ($v = \pi$). The constant K is related to the angular momentum around the x^3-axis. This is not unexpected since due to the rotational symmetry of the torus around the x^3-axis we can expect that the x^3-component of the angular momentum will be conserved. Let us check this, the angular momentum is given by the outer product

$$\mathbf{L} = m\,\mathbf{x} \times \frac{d\mathbf{x}}{dt}$$

The x^3-component ℓ of this quantity is therefore given by

$$\begin{aligned}\ell &= m(x^1\dot{x}^2 - x^2\dot{x}^1) = m(a + b\cos v)\cos u\,[\dot{u}\,(a + b\cos v)\cos u - \dot{v}\,b\sin v\sin u] \\ &\quad - m(a + b\cos v)\sin u\,[-\dot{u}\,(a + b\cos v)\sin u - \dot{v}\,b\sin v\cos u] \\ &= m\dot{u}(a + b\cos v)^2 \end{aligned} \tag{4.44}$$

If we compare this to Eq.(4.43) we see that $K = \ell/m$. The equations of motion therefore indeed tell us that ℓ is a conserved quantity. Inserting Eq.(4.43) into the energy formula (4.42) gives

$$E = \frac{1}{2}m\,b^2\dot{v}^2 + \frac{\ell^2}{2m(a + b\cos v)^2} \tag{4.45}$$

From this equation we see immediately a few useful things. First of all, if the angular momentum is zero $\ell = 0$, then $\dot{v} = (\sqrt{2E/m})/b$ is constant and Eq.(4.43) tells us that $\dot{u} = 0$. We therefore immediately find a solution

$$(u(t), v(t)) = (u_0, \frac{1}{b}\sqrt{\frac{2E}{m}}\,t + v_0) \tag{4.46}$$

where u_0 and v_0 are constants. This geodesic is simply a circle in the plane $u = u_0$ where the particle moves with uniform velocity around the circle. Further insight in the geodesics is obtained by setting $x = bv$ and rewriting Eq.(4.45) as

$$E = \frac{1}{2}m\dot{x}^2 + \frac{\ell^2}{2m(a + b\cos(x/b))^2} \tag{4.47}$$

This is simply the total energy of a single particle moving in an effective potential

$$V(x) = \frac{\ell^2}{2m(a + b\cos(x/b))^2}$$

The corresponding equation of motion

$$m\ddot{x} = -\frac{\partial V}{\partial x} = \frac{\ell^2\sin(x/b)}{m(a + b\cos(x/b))^3} \tag{4.48}$$

is, of course, nothing but Eq.(4.41) after insertion of Eq.(4.43). The shape of the effective potential $V(x)$ is drawn in Fig.4.3

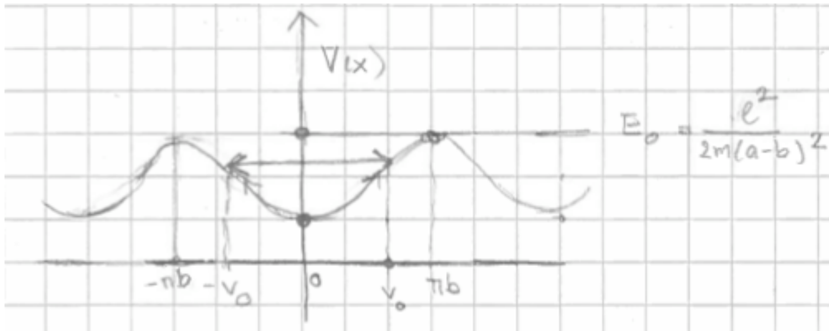

Figure 4.3: Effective potential for the motion in the v-direction on the torus.

This potential is periodic with maxima and minima. One simple solution to the Eq.(4.48) is when the particle is at rest at the minimum or maximum of the potential at $x = 0$ and $x = b\pi$, or equivalently $v = 0$ and $v = \pi$. With help of Eq.(4.43) we find two other geodesics given by

$$(u_1(t), v_1(t)) = (\frac{\ell\, t}{2m(a+b)^2} + u_0, 0)$$
$$(u_2(t), v_2(t)) = (\frac{\ell\, t}{2m(a-b)^2} + u_0, \pi)$$

which describe the uniform motion of the particle along the larger outer or smaller inner circle of the torus in the plane $x^3 = 0$. The other geodesics are not so easy to obtain in closed form but we can easily deduce their properties from Fig.4.3. Let us denote the values of the maxima of the potential by

$$E_0 = \frac{\ell^2}{2m(a-b)^2}$$

If the energy of the effective particle is smaller than E_0 it will oscillate back and forth in the potential. This means that $v(t)$ oscillates between values $-v_0$ and v_0 with $v_0 < \pi$. The particle moves on the torus without ever making a full twist around the tube that forms the torus. If the energy of the particle is larger than E_0 the particle is unbound and $v(t)$ describes a motion in which the particle twists around the tube infinitely many times. An interesting problem is whether some of these orbits are closed orbits. This will be a nice puzzle for you to answer.

4.3 When is the world flat?

We the knowledge gained in the previous section we can now answer the following question. Suppose we are given a metric $g_{ij}(y)$ in coordinates y^j. Then how do we know whether there exists a coordinate transformation to new coordinates $x^k(y)$ such that in these new coordinates

$$g = \sum_k^n dx^k \otimes dx^k \tag{4.49}$$

in other words when can we transform to a Euclidean metric? First of all, we note that a metric of the form (4.49) does not imply that the manifold we consider is equal to $\mathbb{R}^n$. It applies, for instance, also to the cylinder of radius R which we can parametrize as embedded in $\mathbb{R}^3$ as

$$x^1 = R\cos(\phi/R)$$
$$x^2 = R\sin(\phi/R)$$
$$x^3 = z$$

with $z \in \mathbb{R}$ and $\phi \in [0, 2\pi R]$ and which inherits from $\mathbb{R}^3$ induced metric

$$g = d\phi \otimes d\phi + dz \otimes dz \tag{4.50}$$

which is exactly of the form of an Euclidean metric. After some thought this does not surprise us so much since we know that we can easily roll a piece of paper around a cylinder without wrinkling or tearing. This is, for instance, not true when we try to wrap a piece of paper around a sphere or a torus. In some sense these objects are more curved and we suspect that it is not possible to define an Euclidean metric on these surfaces. How can we quantify this? Let us suppose that our given metric $g_{ij}(y)$ can be obtained from an Euclidean metric as we also did in Eq.(4.7). Then, since in the Euclidean coordinate system the Christoffel symbols vanish, we see from Eq.(4.19) that the Christoffel symbols in the y-coordinate system are given by

$$\Gamma^{\gamma}_{\alpha\beta}(y) = \sum_{l}^{n} \frac{\partial^2 x^l}{\partial y^\alpha \partial y^\beta} \frac{\partial y^\gamma}{\partial x^l}$$

or equivalently

$$\frac{\partial^2 x^l}{\partial y^\alpha \partial y^\beta} = \sum_{\gamma}^{n} \Gamma^{\gamma}_{\alpha\beta}(y) \frac{\partial x^l}{\partial y^\gamma}$$

Which can be written as

$$\frac{\partial}{\partial y^\beta}\left(\frac{\partial x^l}{\partial y^\alpha}\right) = \sum_{\gamma}^{n} \Gamma^{\gamma}_{\alpha\beta}(y) \frac{\partial x^l}{\partial y^\gamma} \tag{4.51}$$

We can view this as a set of differential equations for $\partial x^l / \partial y^\alpha$ since the Christoffel symbols are known from our given metric $g_{ij}(y)$. The solutions to these equations determine the desired functions $x^l(y)$ which transform the metric to an Euclidean one. The central question is therefore to determine when the equations (4.51) have a solution. Note further that the index l does not play any special role since all functions x^l satisfy exactly the same equation. Let us denote the vector a and the n vector valued functions f_β by

$$\begin{aligned} a^\alpha &= \frac{\partial x^l}{\partial y^\alpha} \\ f^\alpha_\beta(y, z) &= \sum_{\gamma}^{n} \Gamma^{\gamma}_{\alpha\beta}(y)\, z^\gamma \end{aligned} \tag{4.52}$$

Then the differential equations (4.51) attain the form

$$\frac{\partial a^\alpha}{\partial y^\beta} = f^\alpha_\beta(y, a(y)) \tag{4.53}$$

This is a differential equation which is linear in a. If these equations have a solution then there will be n linearly independent solutions a (corresponding to n different coordinate functions $x^l(y)$). The necessary conditions for the existence of a solution to the Eqs.(4.53) are given by

$$\frac{\partial f^\alpha_\beta}{\partial y^\rho} = \frac{\partial a^\alpha}{\partial y^\rho \partial y^\beta} = \frac{\partial a^\alpha}{\partial y^\beta \partial y^\rho} = \frac{\partial f^\alpha_\rho}{\partial y^\beta}$$

which can also be shown to be sufficient conditions. These integrability conditions can be rewritten as

$$\begin{aligned} 0 &= \frac{\partial f^\alpha_\beta}{\partial y^\rho}(y, a(y)) - \frac{\partial f^\alpha_\rho}{\partial y^\beta}(y, a(y)) = \frac{\partial f^\alpha_\beta}{\partial y^\rho}|_z - \frac{\partial f^\alpha_\rho}{\partial y^\beta}|_z + \sum_{\mu}^{n} \frac{\partial f^\alpha_\beta}{\partial z^\mu}|_y \frac{\partial a^\mu}{\partial y^\rho} - \sum_{\mu}^{n} \frac{\partial f^\alpha_\rho}{\partial z^\mu}|_y \frac{\partial a^\mu}{\partial y^\beta} \\ &= \frac{\partial f^\alpha_\beta}{\partial y^\rho}|_z - \frac{\partial f^\alpha_\rho}{\partial y^\beta}|_z + \sum_{\mu}^{n} \frac{\partial f^\alpha_\beta}{\partial z^\mu}|_y f^\mu_\rho - \sum_{\mu}^{n} \frac{\partial f^\alpha_\rho}{\partial z^\mu}|_y f^\mu_\beta \end{aligned} \tag{4.54}$$

If we insert the explicit form of the functions f from Eq.(4.52) we obtain

$$0 = \sum_{\gamma}^{n} \frac{\partial \Gamma^{\gamma}_{\alpha\beta}}{\partial y^{\rho}} z^{\gamma} - \sum_{\gamma}^{n} \frac{\partial \Gamma^{\gamma}_{\alpha\rho}}{\partial y^{\beta}} z^{\gamma} + \sum_{\mu}^{n} \Gamma^{\mu}_{\alpha\beta} \sum_{\gamma}^{n} \Gamma^{\gamma}_{\mu\rho}\, z^{\gamma} - \sum_{\mu}^{n} \Gamma^{\mu}_{\alpha\rho} \sum_{\gamma}^{n} \Gamma^{\gamma}_{\mu\beta}\, z^{\gamma}$$

Since there are n-independent solutions to Eq.(4.51) this condition must be valid for all $z = (z^1, \ldots, z^n)$ and we obtain

$$R^{\gamma}_{\alpha\rho\beta} = 0 \tag{4.55}$$

where we defined

$$R^{\gamma}_{\alpha\rho\beta} = \frac{\partial \Gamma^{\gamma}_{\alpha\beta}}{\partial y^{\rho}} - \frac{\partial \Gamma^{\gamma}_{\alpha\rho}}{\partial y^{\beta}} + \sum_{\mu}^{n} (\Gamma^{\mu}_{\alpha\beta} \Gamma^{\gamma}_{\mu\rho} - \Gamma^{\mu}_{\alpha\rho} \Gamma^{\gamma}_{\mu\beta}) \tag{4.56}$$

The vanishing of these coefficients imply the existence of a coordinate transformation to an Euclidean metric. Since the condition should be valid in any coordinate system y that we started with we already can expect these coefficients to be coefficients of a tensor. This is indeed true, the corresponding tensor is known as the Riemann tensor which is a mixed tensor of type $(3, 1)$. We have not shown that it is a tensor, but if you have a lot of paper and patience you can check from Eq.(4.19) for the transformation law of the Christoffel symbols that under coordinate transformations it transforms correctly as

$$R'^{\alpha}_{\beta\gamma\delta}(y) = \sum_{i,j,k,l}^{n} R^{i}_{jkl}(x) \frac{\partial x^j}{\partial y^{\beta}} \frac{\partial x^k}{\partial y^{\gamma}} \frac{\partial x^l}{\partial y^{\delta}} \frac{\partial y^{\alpha}}{\partial x^i} \tag{4.57}$$

The non-vanishing of R^i_{jkl} implies that we our metric corresponds to that of a curved manifold. The tensor has n^4 components so to check that all of them vanish maybe a lot of work. However, not all components are independent. From the definition (4.56) we see that

$$R^i_{jkl} = -R^i_{jlk} \tag{4.58}$$

$$0 = R^i_{jkl} + R^i_{klj} + R^i_{ljk} \tag{4.59}$$

Another way to reduce the number of indices is to perform a contraction of the upper index and a lower one. In fact, there is only one independent tensor that can be obtained in this way. To see this we need a few more hidden symmetries which become apparent by using the explicit form of the Christoffel symbols and lowering the upper index. This gives the tensor

$$R_{ijkl} = \sum_{m}^{n} g_{i\gamma} R^{\gamma}_{jkl} \tag{4.60}$$

The vanishing of this tensor is equivalent to the vanishing of the Riemann tensor, so we can also use this tensor to find out if our metric describes a curved space. With help of the relation

$$\sum_{\gamma}^{n} g_{i\gamma} \frac{\partial \Gamma^{\gamma}_{jl}}{\partial y^k} = \frac{\partial}{\partial y^k} \left(\sum_{\gamma}^{n} g_{i\gamma} \Gamma^{\gamma}_{jl} \right) - \sum_{\gamma}^{n} \Gamma^{\gamma}_{jl} \frac{\partial g_{i\gamma}}{\partial y^k} = \frac{\partial [jl, i]}{\partial y^k} - \sum_{\gamma}^{n} \Gamma^{\gamma}_{jl} ([ik, \gamma] + [\gamma k, i])$$

as well as the explicit form of the Christoffel symbols of the first kind of Eq.(4.12) we find from Eq.(4.60) that

$$\begin{aligned} R_{ijkl} =& \frac{1}{2} \left(\frac{\partial^2 g_{il}}{\partial y^j \partial y^k} + \frac{\partial^2 g_{jk}}{\partial y^i \partial y^l} - \frac{\partial^2 g_{ik}}{\partial y^j \partial y^l} - \frac{\partial^2 g_{jl}}{\partial y^i \partial y^k} \right) \\ &+ \sum_{\alpha,\beta}^{n} g^{\alpha\beta} ([jk, \alpha][il, \beta] - [ik, \alpha][jl, \beta]) \end{aligned} \tag{4.61}$$

This equation reveals the further symmetries

$$R_{ijkl} = -R_{ijlk} = -R_{jikl} = R_{klij} \tag{4.62}$$

In particular if follows from the anti-symmetry in the first two indices that the following contraction vanishes,

$$\sum_i R^i_{ijk} = \sum_{i,\gamma} g^{im} R_{mijk} = 0$$

since the metric tensor g^{im} is symmetric in i and m. Then taking $i = j$ in Eq.(4.59) and summing over i then gives

$$\sum_i^n R^i_{lik} = -\sum_i^n R^i_{kli}$$

There is therefore only one possible independent contraction of the Riemann tensor, which is called the Ricci tensor R_{lk} defined by

$$R_{lk} = \sum_i^n R^i_{lik} = -\sum_i^n R^i_{kli} = \sum_i^n R^i_{kil} = R_{kl} \tag{4.63}$$

where in the last step we used Eq.(4.58) to switch the last two lower indices. The Ricci tensor is therefore a symmetric tensor. Finally, there is the possibility to define a scalar called the Ricci scalar by

$$R = \sum_{i,j}^n g^{ij} R_{ij} \tag{4.64}$$

The Ricci and Riemann tensors can be used to characterize curved spaces. If the Ricci tensor or scalar does not vanish then the space does not admit an Euclidean metric. It is possible that the Ricci tensor vanishes while the Riemann tensor does not[1] so the most robust characterization of a curved space is given by the Riemann tensor.

Let us now given an example to illustrate our derivations. We calculate the Riemann tensor for the torus. Because of the symmetries (4.62) and the fact that we only have two indices we find that for two-dimensional surfaces there is only one non-vanishing independent component, which for the torus in u and v coordinates is given by

$$R_{uvuv} = -R_{vuuv} = R_{vuvu} = -R_{uvvu}$$

We can calculate R_{uvuv} directly from Eq.(4.61). The metric is given in Eq.(3.141) and the relevant Christoffel symbols were already calculated in Eqs.(4.26) and (4.27). Since $g_{uv} = 0$ and g_{vv} is constant we have from (4.61)

$$R_{uvuv} = -\frac{1}{2}\frac{\partial^2 g_{uu}}{\partial v^2} + \sum_{\alpha,\beta} g^{\alpha\beta}([vu,\alpha][uv,\beta] - [uu,\alpha][vv,\beta])$$

Since the only non-vanishing Christoffel symbols of are $[uv,u] = [vu,u]$ and $[uu,v]$ we find

$$R_{uvuv} = -\frac{1}{2}\frac{\partial^2}{\partial v^2}(a + b\cos v)^2 + g^{uu}[uv,u][vu,u] = b\cos v(a + b\cos v)$$

This immediately gives the non-vanishing coefficients of the Riemann tensor

$$R^u_{vuv} = g^{uu} R_{uvuv} = \frac{b\cos v}{a + b\cos v} = -R^u_{vvu}$$
$$R^v_{uvu} = g^{vv} R_{vuvu} = \frac{1}{b}\cos v(a + b\cos v) = -R^v_{uuv}$$

[1]In general relativity a well-known example is that of the Schwarzschild solution of the black hole for which the Ricci tensor vanishes but the Riemann tensor does not.

For the Ricci tensor R_{ik} we find that the only non-vanishing elements are

$$R_{uu} = R^v_{uvu} = \frac{1}{b}\cos v(a + b\cos v) \quad , \quad R_{vv} = R^u_{vuv} = \frac{b\cos v}{a + b\cos v}$$

whereas the Ricci scalar is given by

$$R = g^{uu}R_{uu} + g^{vv}R_{vv} = \frac{2\cos v}{b(a + b\cos v)}$$

In any case, we see that the Riemann tensor for the torus does not vanish and therefore there does not exist a coordinate transformation that would make its metric Euclidean.
As a final note we remark that the Ricci tensor plays an important role in the general theory of relativity where it is linearly related to the energy-momentum tensor about which we will hear more later. In any case, the physical picture is that energy and mass distributions determine the form of R_{ik} from which we can calculate the metric (up to a coordinate transformation). In the language of the examples of embedded surfaces we have been using, the energy and momentum distribution determines the shape of the manifold on which the particle is moving. However, this shape is in general not static but changing in time, for instance due to gravitational waves, which in our picture describe moving ripples along the surface of the manifold. In any case, an important thing to remember is that the shape of the manifold is completely independent of the type of coordinate system one wants to employ on it and therefore the physical laws should be given in a way that does not depend on the coordinate system.

Chapter 5

The covariant derivative

We will discuss how to differentiate vectors and tensors in general coordinates. We start out by discussing differentiation of vector fields in flat space. Then we discuss how to differentiate vector fields on a surface embedded in three-dimensional space. This motivates our final general definitions of the covariant derivative of vector and tensor fields which do not rely on any embedding in a higher dimensional flat space. We further discuss the geometric meaning of the vanishing of the covariant derivative of the metric tensor.

5.1 Differentiating vector fields along curves

In the previous Chapter we saw that constrained particle motion gives a nice geometric and easily imaginable picture of motion in general coordinate systems. We start by exploring this picture a bit further by studying how vector fields change along particle trajectories. Since this is a more complicated concept than the change of scalar fields we introduce the concept in three steps. We first consider vector fields in flat space, then we consider vector fields along surfaces embedded in three-dimensional flat space and finally discuss the general case of curved spaces. Let us start with the case of vector fields in flat space.

Imagine a flat two-dimensional plane in which at each point we assign a vector w. This could, for instance, represent a wind flow over the surface of the plane or the water flow in a river. We assume that the vector field is static and that the vectors only depend on their spatial position. We can now imagine that we can travel along a path $c(t)$ parametrized by time t in the plane and at each point measure the two components (w^1, w^2) of the vector. To do this we can construct a Cartesian frame with coordinates (x^1, x^2) and orthogonal basis vectors (e_1, e_2) as displayed in Fig.5.1a. The path has coordinates $c(t) = (x^1(t), x^2(t))$ and at each point along the path we register the two components $(w^1(t), w^2(t))$ with respect to the orthonormal basis (e_1, e_2).

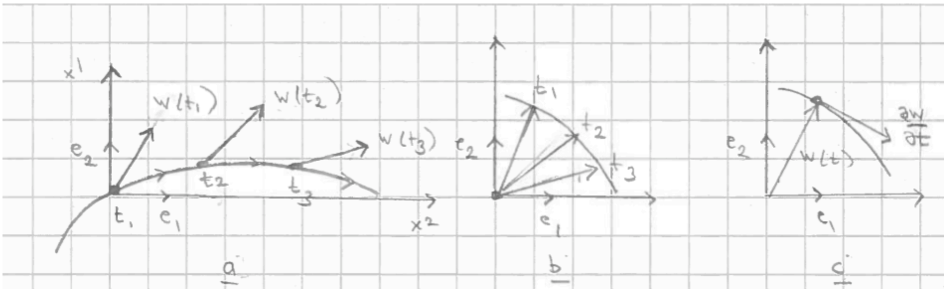

Figure 5.1: A vector field w along a path $c(t)$ (a). We rigidly translated the vectors to the same point (b). These endpoints trace out a path in time and the time derivative $\partial w/\partial t$ is a tangent vector to this path (c).

In Fig. 5.1a we display three of such vectors registered at three different times. To compare these vectors at different locations we can translate each of the vectors rigidly along the coordinate axes to the same point and compare them in a single graph. This is done in Fig.5.1b. In this graph the endpoints of the vectors $w(t)$ trace out a path in time. Then we can calculate the derivative

$$\frac{\partial w}{\partial t} = (\frac{\partial w^1}{\partial t}, \frac{\partial w^2}{\partial t}) \tag{5.1}$$

which is the tangent vector to the path traced out by the endpoints of the vectors $w(t)$ as displayed in Fig. 5.1c. We can express this more explicitly in coordinates as

$$\frac{\partial w^j}{\partial t} = \sum_{k=1}^{2} \frac{\partial w^j}{\partial x^k}\frac{\partial x^k}{\partial t} = \sum_{k=1}^{2} \frac{\partial w^j}{\partial x^k}\,\dot{c}^k(t) \tag{5.2}$$

where we introduced the notation

$$\frac{\partial c(t)}{\partial t} = (\frac{\partial x^1}{\partial t}, \frac{\partial x^1}{\partial t}) = (\dot{c}^1, \dot{c}^2) = \dot{c}(t) \tag{5.3}$$

which is just the tangent vector along the path $c(t)$. The expression (5.2) represents the change of the vector field w if we move in the direction of the vector $\dot{c}$. We will denote this new vector by $\nabla_{\dot{c}}\, w$. In components

$$(\nabla_{\dot{c}}\, w)^j = \sum_{k=1}^{2} \frac{\partial w^j}{\partial x^k}\,\dot{c}^k(t) \tag{5.4}$$

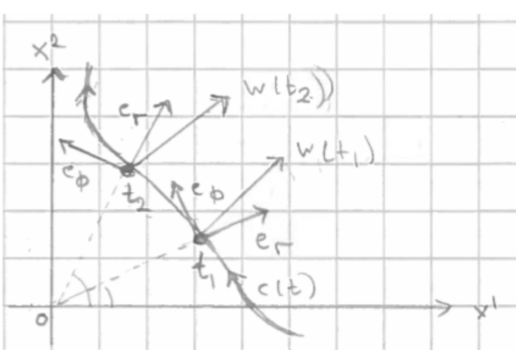

Figure 5.2: The vector field along a path $c(t)$ from as viewed in a polar coordinate system.

Let us now see what happens in a different coordinate frame, such as the polar coordinate system discussed in Fig. 3.2 of Chapter 3. In Fig. 5.2 we display a vector field along a path $c(t)$ in a polar coordinate system. Now we see that the basis vectors e_r and e_ϕ rotate as we move along the path, which affects the rate of change of component projections of the vectors w as we move along the path. For instance, the vectors $w(t_1)$ and $w(t_2)$ in the figure have very different components (w^r, w^ϕ) with respect to the polar basis vectors, although the vectors are almost parallel from the viewpoint of the Cartesian coordinate system. To calculate the vector $\nabla_{\dot{c}}\, w$ in the polar coordinate system we therefore also have to take into account the change in the basis vectors (e_r, e_ϕ) when we move along the path $c(t)$. Since we already calculated this quantity in Eq.(5.2) we can find the result in any other coordinate system by transforming this result to different coordinates. Let us call these new coordinates (y^1, y^2). in these coordinates we have

$$\frac{\partial w^j}{\partial t} = \frac{\partial}{\partial t}\left(\sum_k w'^k \frac{\partial x^j}{\partial y^k}\right) = \sum_k \frac{\partial w'^k}{\partial t}\frac{\partial x^j}{\partial y^k} + \sum_{k,l} w'^k \frac{\partial^2 x^j}{\partial y^k \partial y^l}\frac{\partial y^l}{\partial t} \tag{5.5}$$

The last term can be expressed in terms of the Christoffel symbols of the new coordinate system. Now in the Cartesian coordinate system the metric tensor is simply given by $g_{ij} = \delta_{ij}$ and therefore the Christoffel symbols vanish in this system. Therefore the transformation law Eq.(4.19) tells us that the Christoffel symbols in the new coordinate frame are given by

$$\Gamma'^{\,p}_{kl} = \sum_q \frac{\partial^2 x^q}{\partial y^k \partial y^l}\frac{\partial y^p}{\partial x^q}$$

or equivalently

$$\frac{\partial^2 x^j}{\partial y^k \partial y^l} = \sum_p \Gamma'^{\,p}_{kl}\frac{\partial x^j}{\partial y^p}$$

If we insert this into Eq.(5.5) we find

$$\begin{aligned}(\nabla_{\dot{c}}\, w)^j = \frac{\partial w^j}{\partial t} &= \sum_p \frac{\partial w'^p}{\partial t}\frac{\partial x^j}{\partial y^p} + \sum_{k,l,p} w'^k \Gamma'^{\,p}_{kl}\frac{\partial x^j}{\partial y^p}\frac{\partial y^l}{\partial t} \\ &= \sum_p \frac{\partial x^j}{\partial y^p}\left(\frac{\partial w'^p}{\partial t} + \sum_{k,l} w'^k \Gamma'^{\,p}_{kl}\frac{\partial y^l}{\partial t}\right) = \sum_p \frac{\partial x^j}{\partial y^p}(\nabla_{\dot{c}}\, w)'^{\,p}\end{aligned} \tag{5.6}$$

where we denote

$$\begin{aligned}(\nabla_{\dot{c}}\, w)'^{\,p} = \frac{\partial w'^p}{\partial t} + \sum_{k,l} w'^k \Gamma'^{\,p}_{kl}\frac{\partial y^l}{\partial t} &= \sum_l \left(\frac{\partial w'^p}{\partial y^l} + \sum_k w'^k \Gamma'^{\,p}_{kl}\right)\frac{\partial y^l}{\partial t} \\ &= \sum_l \left(\frac{\partial w'^p}{\partial y^l} + \sum_k w'^k \Gamma'^{\,p}_{kl}\right)\dot{c}'^{\,l}\end{aligned} \tag{5.7}$$

where $c'(t) = (y^1(t), y^2(t))$ is the path $c(t)$ in the y-coordinate system. Equation (5.7) is the main result of our derivation. It represents a generalization to general coordinates of the vector Eq.(5.4) in the Cartesian coordinate system. Equation (5.6) shows that this quantity indeed transforms as a vector under coordinate transformations. We will derive this in more detail below as well.

Now we consider a slightly more difficult case, namely that of a vector field along a surface. This will at the same time help to get a very nice geometrical insight into the expression (5.7). We consider a vector field along a curve $c(t)$ that describes the motion of a particle in a surface parametrized by coordinates (u^1, u^2), i.e. we have $\mathbf{x} = (x^1, x^2, x^3)$ with $x^j = x^j(u^1, u^2)$. We consider again a vector field w which is tangent to the surface, for instance, a vector field that describes a wind flow or temperature gradient on the surface of the Earth. Pictorially we have

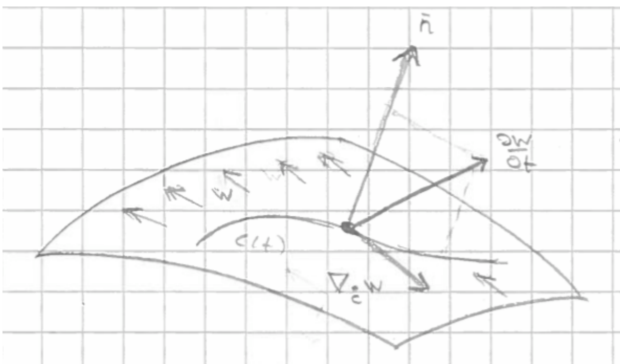

Figure 5.3: A tangent vector field w along a surface. The vector $\nabla_{\dot{c}}\, w$ is defined as the tangential component of $\partial w/\partial t$.

The curve has the form $c(t) = (u^1(t), u^2(t))$ in terms of surface coordinates. The velocity vector

$$\dot{c}(t) = \dot{u}^1(t)\frac{\partial}{\partial u^1} + \dot{u}^2(t)\frac{\partial}{\partial u^2}$$

has, when mapped to the surrounding three-dimensional space using an embedding mapping the form

$$\dot{c}(t) = \dot{u}^1(t)\frac{\partial \mathbf{x}}{\partial u^1} + \dot{u}^2(t)\frac{\partial \mathbf{x}}{\partial u^2}$$

with respect to Cartesian basis vectors in $\mathbb{R}^3$, where $\partial\mathbf{x}/\partial u^i$ are the tangents to the coordinate curves on the surface. Since the vector field w is tangential it can be expressed as

$$w = w^1\frac{\partial \mathbf{x}}{\partial u^1} + w^2\frac{\partial \mathbf{x}}{\partial u^2} \tag{5.8}$$

We want to see how this vector field changes when we move along the curve $c(t)$. As in the case of the two-dimensional plane (see Fig.5.1b), we can move the vectors $w(t)$ at different points to the same point in space by rigid translations along the coordinate axes of the surrounding three-dimensional space. The end points of these vectors trace out a curve in three-dimensional space and we can calculate the tangent to this curve. We have

$$\begin{aligned}\frac{\partial w}{\partial t} &= \frac{\partial}{\partial t}\left(\sum_i w^i \frac{\partial \mathbf{x}}{\partial u^i}\right) = \sum_i \frac{\partial w^i}{\partial t}\frac{\partial \mathbf{x}}{\partial u^i} + \sum_{i,j} w^i \frac{\partial^2 \mathbf{x}}{\partial u^i \partial u^j}\dot{u}^j \\ &= \sum_{i,j}\left(\frac{\partial w^i}{\partial u^j}\frac{\partial \mathbf{x}}{\partial u^i}\dot{u}^j + w^i\frac{\partial^2 \mathbf{x}}{\partial u^i \partial u^j}\dot{u}^j\right)\end{aligned} \tag{5.9}$$

The first term in Eq.(5.9) is a linear combination of $\partial\mathbf{x}/\partial u^i$ and therefore represents a vector in the tangent plane to the surface. The last term, however, has in general also components pointing out of the surface. This is not difficult to understand physically. In the special case that we take $w(c(t)) = \dot{c}(t)$ the vector $\partial w/\partial t$ would represent the the acceleration of the particle which in general also has a normal component that wants to push the particle outside of the surface. What we now can do is to calculate the component of $\partial w/\partial t$ parallel to the surface which we will call $\nabla_{\dot c}\, w$ (the relation to Eq.(5.7) will become clear soon). It is given by

$$\nabla_{\dot c}\, w = \frac{\partial w}{\partial t} - \mathbf{n}\,\langle \frac{\partial w}{\partial t}, \mathbf{n}\rangle \tag{5.10}$$

where $\mathbf{n}$ is a unit normal vector to the surface and we used the standard Euclidean inner product. So $\langle \mathbf{n}, \nabla_{\dot c}\, w\rangle = 0$, which means that we projected $\partial w/\partial t$ back into the tangent plane. The notation $\nabla_{\dot c}\, w$ is therefore a short notation for the rate of change of the vector field w within the tangent plane in the direction of $\dot c$. The mapping

$$w \to \nabla_{\dot c}\, w$$

for a given tangent vector $\dot c$ to the path $c(t)$ in a point p maps tangent vectors in T_pM to new tangent vectors of T_pM. Since the definition (5.10) is a geometric one, independent of the surface parametrization, $\nabla_{\dot c}\, w$ is a coordinate invariant object and we expect it to transform like a vector under change of surface coordinates. We will see that this is indeed the case.
First of all, since $\nabla_{\dot c}\, w$ lies in the tangent plane we can expand it in terms of $\partial\mathbf{x}/\partial u^i$ as

$$\nabla_{\dot c}\, w = \sum_i (\nabla_{\dot c}\, w)^i \frac{\partial \mathbf{x}}{\partial u^i} \tag{5.11}$$

To find a more explicit expression for the coefficients $(\nabla_{\dot c}\, w)^i$ we take the inner product on both sides of this expression with $\partial\mathbf{x}/\partial u^j$. This gives

$$\langle \nabla_{\dot c}\, w, \frac{\partial \mathbf{x}}{\partial u^j}\rangle = \sum_i (\nabla_{\dot c}\, w)^i \langle \frac{\partial \mathbf{x}}{\partial u^i}, \frac{\partial \mathbf{x}}{\partial u^j}\rangle = \sum_i (\nabla_{\dot c}\, w)^i\, g_{ij} \tag{5.12}$$

where we used Eq.(3.140), i.e.

$$g_{ij} = \langle \frac{\partial \mathbf{x}}{\partial u^i}, \frac{\partial \mathbf{x}}{\partial u^j}\rangle$$

Consequently we find that

$$(\nabla_{\dot c}\, w)^l = \sum_k g^{lk}\, \langle \nabla_{\dot c}\, w, \frac{\partial \mathbf{x}}{\partial u^k}\rangle \tag{5.13}$$

from Eqs.(5.10) and (5.9) we then see that

$$\begin{aligned}\langle \nabla_{\dot c}\, w, \frac{\partial \mathbf{x}}{\partial u^k}\rangle = \langle \frac{\partial w}{\partial t}, \frac{\partial \mathbf{x}}{\partial u^k}\rangle &= \sum_{i,j} \left(\frac{\partial w^i}{\partial u^j}\dot u^j \langle \frac{\partial \mathbf{x}}{\partial u^i}, \frac{\partial \mathbf{x}}{\partial u^k}\rangle + w^i \dot u^j \langle \frac{\partial^2 \mathbf{x}}{\partial u^i \partial u^j}, \frac{\partial \mathbf{x}}{\partial u^k}\rangle \right) \\ &= \sum_{i,j} \left(\frac{\partial w^i}{\partial u^j}\dot u^j\, g_{ik} + w^i \dot u^j \langle \frac{\partial^2 \mathbf{x}}{\partial u^i \partial u^j}, \frac{\partial \mathbf{x}}{\partial u^k}\rangle \right)\end{aligned}$$

Inserting this back into Eq.(5.13) then gives

$$(\nabla_{\dot c}\, w)^l = \sum_j \frac{\partial w^l}{\partial u^j}\dot u^j + \sum_{i,j,k} w^i \dot u^j g^{lk} \langle \frac{\partial^2 \mathbf{x}}{\partial u^i \partial u^j}, \frac{\partial \mathbf{x}}{\partial u^k}\rangle \tag{5.14}$$

We now only need a more explicit expression for the last term in this equation. We can do this by calculating the derivatives of the coefficients of the metric tensor. using the short notation

$$\langle ij, k\rangle = \langle \frac{\partial^2 \mathbf{x}}{\partial u^i \partial u^j}, \frac{\partial \mathbf{x}}{\partial u^k}\rangle \tag{5.15}$$

we have

$$\begin{aligned}
\frac{\partial g_{ik}}{\partial u^j} &= \frac{\partial}{\partial u^j}\langle \frac{\partial \mathbf{x}}{\partial u^i}, \frac{\partial \mathbf{x}}{\partial u^k}\rangle = \langle ij,k\rangle + \langle jk,i\rangle \\
\frac{\partial g_{jk}}{\partial u^i} &= \frac{\partial}{\partial u^i}\langle \frac{\partial \mathbf{x}}{\partial u^j}, \frac{\partial \mathbf{x}}{\partial u^k}\rangle = \langle ij,k\rangle + \langle ik,j\rangle \\
\frac{\partial g_{ij}}{\partial u^k} &= \frac{\partial}{\partial u^k}\langle \frac{\partial \mathbf{x}}{\partial u^i}, \frac{\partial \mathbf{x}}{\partial u^j}\rangle = \langle ik,j\rangle + \langle jk,i\rangle
\end{aligned}$$

and we therefore see that

$$\langle ij,k\rangle = \frac{1}{2}\left(\frac{\partial g_{ik}}{\partial u^j} + \frac{\partial g_{jk}}{\partial u^i} - \frac{\partial g_{ij}}{\partial u^k}\right)$$

Comparison with Eq.(4.12) then tells us that $\langle ij,k\rangle = [ij,k]$ and we see that they are equal to the Christoffel symbols of the first kind. Then Eq.(5.14) becomes

$$\begin{aligned}
(\nabla_{\dot{c}}\, w)^l &= \sum_j \frac{\partial w^l}{\partial u^j}\dot{u}^j + \sum_{i,j,k} w^i \dot{u}^j g^{lk}[ij,k] \\
&= \sum_j \left(\frac{\partial w^l}{\partial u^j} + \sum_i w^i \Gamma^l_{ij}\right)\dot{u}^j = \sum_j w^l_{;j}\dot{u}^j \qquad (5.16)
\end{aligned}$$

where we defined

$$w^l_{;j} = \frac{\partial w^l}{\partial u^j} + \sum_i w^i \Gamma^l_{ij} \qquad (5.17)$$

It remains to show that $\nabla_{\dot{c}}\, w$ actually transforms as a vector. We start by showing that the components $w^i_{;j}$ transform as the components of a mixed tensor of type $(1,1)$, i.e.

$$w^i_{;j} = \sum_{\alpha,\beta} w'^{\alpha}_{;\beta}\, \frac{\partial u^i}{\partial u'^{\alpha}} \frac{\partial u'^{\beta}}{\partial u^j}$$

The validity of this equation only depends on the transformation law of the Christoffel symbols and hence is valid for any n-dimensional manifold. Instead of (u^1,u^2) we therefore use $(x^1,\ldots,x^n)$ as coordinates. Let us then write

$$w^l_{;j} = \frac{\partial w^l}{\partial x^j} + \sum_i w^i \Gamma^l_{ij} \qquad (5.18)$$

then in new coordinates y^j we have using the transformation law Eq.(4.19) for the Christoffel symbols that

$$\begin{aligned}
w^l_{;j} &= \frac{\partial}{\partial x^j}\left(\sum_k w'^k \frac{\partial x^l}{\partial y^k}\right) + \sum_{i,k} w'^k \frac{\partial x^i}{\partial y^k}\left(\sum_{\alpha,\beta,\gamma} \Gamma'^{\gamma}_{\alpha\beta} \frac{\partial y^\alpha}{\partial x^i}\frac{\partial y^\beta}{\partial x^j}\frac{\partial x^l}{\partial y^\gamma} + \sum_p \frac{\partial^2 y^p}{\partial x^i \partial x^j}\frac{\partial x^l}{\partial y^p}\right) \\
&= \sum_{p,k} \frac{\partial w'^k}{\partial y^p}\frac{\partial y^p}{\partial x^j}\frac{\partial x^l}{\partial y^k} + \sum_{p,k} w'^k \frac{\partial^2 x^l}{\partial y^k \partial y^p}\frac{\partial y^p}{\partial x^j} + \sum_{\alpha,\beta,\gamma} w'^\alpha \Gamma'^{\gamma}_{\alpha\beta}\frac{\partial y^\beta}{\partial x^j}\frac{\partial x^l}{\partial y^\gamma} + \sum_{i,p,k} w'^k \frac{\partial x^i}{\partial y^k}\frac{\partial^2 y^p}{\partial x^i \partial x^j}\frac{\partial x^l}{\partial y^p} \\
&= \sum_{p,k}\left(\frac{\partial w'^k}{\partial y^p} + \sum_q w'^q \Gamma'^k_{qp}\right)\frac{\partial y^p}{\partial x^j}\frac{\partial x^l}{\partial y^k} + \sum_k w'^k \left(\sum_p \frac{\partial^2 x^l}{\partial y^k \partial y^p}\frac{\partial y^p}{\partial x^j} + \sum_{i,p}\frac{\partial x^i}{\partial y^k}\frac{\partial^2 y^p}{\partial x^i \partial x^j}\frac{\partial x^l}{\partial y^p}\right)
\end{aligned}$$

Now the last term in this equation vanishes since it represents the derivative of

$$\delta^l_j = \frac{\partial x^l}{\partial x^j} = \sum_p \frac{\partial x^l}{\partial y^p}\frac{\partial y^p}{\partial x^j}$$

with respect to y^k which is zero, and we therefore obtain

$$w^l_{;j} = \sum_{p,k} w'^{\,k}_{;p}\frac{\partial y^p}{\partial x^j}\frac{\partial x^l}{\partial y^k} \tag{5.19}$$

which is exactly what we wanted to prove. If we now use Eq.(5.19) in Eq.(5.16) then we see that

$$\begin{aligned}(\nabla_{\dot c}\, w)^l &= \sum_j v^l_{;j}\dot u^j = \sum_{j,\alpha,\beta,\gamma} w'^{\alpha}_{;\beta}\frac{\partial u^l}{\partial u'^{\alpha}}\frac{\partial u'^{\beta}}{\partial u^j}\frac{\partial u^j}{\partial u'^{\gamma}}\dot u'^{\gamma} = \sum_{\alpha,\beta} w'^{\alpha}_{;\beta}\dot u'^{\beta}\frac{\partial u^l}{\partial u'^{\alpha}} \\ &= \sum_\alpha (\nabla_{\dot c}\, w)'^{\,\alpha}\frac{\partial u^l}{\partial u'^{\alpha}} \end{aligned} \tag{5.20}$$

which is exactly the transformation law of a vector. If we now go back to the vector $\nabla_{\dot c}\, w$ of Eq.(5.11) rather than its components we see that

$$\nabla_{\dot c}\, w = \sum_i (\nabla_{\dot c}\, w)^i \frac{\partial \mathbf{x}}{\partial u^i} = \sum_{i,\alpha}(\nabla_{\dot c}\, w)'^{\alpha}\frac{\partial u^i}{\partial u'^{\alpha}}\frac{\partial \mathbf{x}}{\partial u^i} = \sum_\alpha (\nabla_{\dot c}\, w)'^{\alpha}\frac{\partial \mathbf{x}}{\partial u'^{\alpha}} \tag{5.21}$$

In this equation we wrote $\nabla_{\dot c}\, w$ as a three-dimensional vector, but it is clear that the equation is just as valid when we write it as

$$\nabla_{\dot c}\, w = \sum_i (\nabla_{\dot c}\, w)^i \frac{\partial}{\partial u^i} = \sum_\alpha (\nabla_{\dot c}\, w)'^{\alpha}\frac{\partial}{\partial u'^{\alpha}}$$

in which the vectors live on a two-dimensional manifold parametrized by (u^1, u^2). This is not surprising since

$$\frac{\partial \mathbf{x}}{\partial u^j} = \sum_{k=1}^3 \frac{\partial x^j}{\partial u^j} e_k = \sum_{k=1}^3 \frac{\partial x^j}{\partial u^j}\frac{\partial}{\partial x^k} = i_*(\frac{\partial}{\partial u^j})$$

i.e. Eq.(5.21) is simply the push forward of Eq.(5.22) by the embedding mapping i of the surface in the three-dimensional space. Since Eq.(5.22) is generally valid, what we could have done, is to skip the whole geometric motivation based on the embedded surface and state the following definition of a directional derivative on a general manifold. Given a n-dimensional vector field

$$v = \sum_j^n v^j \frac{\partial}{\partial x^j}$$

and given another vector field

$$w = \sum_j^n w^j \frac{\partial}{\partial x^j}$$

there is a new vector field $\nabla_v\, w$, called the derivative of w in the direction of v given by

$$\nabla_v\, w = \sum_j^n (\nabla_v\, w)^j \frac{\partial}{\partial x^j}$$

with components

$$(\nabla_v w)^k = \sum_j^n \left(\frac{\partial w^k}{\partial x^j} + \sum_i^n \Gamma^k_{ji} w^i \right) v^j \tag{5.22}$$

The proof that $\nabla_v w$ is a vector field is simple. Just do the coordinate transformation and use the transformation properties of the Christoffel symbols as well as the vectors v and w. Although the definition (5.22) is easy to understand algebraically and the proof straightforward the geometric meaning would have been hard to grasp if we had not given the motivation on the previous pages.

5.2 A fancy definition

Now that we know that Eq.(5.22) gives the components of a vector we can reverse the logic and give a "clean" coordinate independent definition of $\nabla_v w$ on a manifold M. Let us start by deriving some properties of $\nabla_v w$. First of all, it is immediately clear from Eq.(5.22) that

$$\nabla_{v_1+v_2} w = \nabla_{v_1} w + \nabla_{v_2} w \tag{5.23}$$
$$\nabla_v (w_1 + w_2) = \nabla_v w_1 + \nabla_v w_2 \tag{5.24}$$
$$\nabla_{f\,v} w = f \nabla_v w \tag{5.25}$$
$$\nabla_v (f\, w) = f\, \nabla_v w + v(f)\, w \tag{5.26}$$

where $f : M \to \mathbb{R}$ is a function on the manifold and v, v_i, w, w_i vector fields defined on M. This brings us to the following definition:
A *connection* or *covariant derivative* on a manifold M is a function ∇ which assigns a vector field $\nabla_v w$ to any two vector fields v and w and which satisfies Eqs.(5.23)-(5.26).
Note that this definition does not require that we have defined a metric on the manifold M. Given the properties (5.23)-(5.26) it is not difficult to derive an explicit form of $\nabla_v w$. We have

$$\begin{aligned}
\nabla_v w = \nabla_v \left(\sum_k^n w^k \frac{\partial}{\partial x^k} \right) &= \sum_k^n \nabla_v \left(w^k \frac{\partial}{\partial x^k} \right) = \sum_k^n \left(w^k \nabla_v \left(\frac{\partial}{\partial x^k} \right) + v(w^k) \frac{\partial}{\partial x^k} \right) \\
&= \sum_{j,k}^n v^j \frac{\partial w^k}{\partial x^j} \frac{\partial}{\partial x^k} + \sum_{j,k}^n w^k v^j \nabla_{\frac{\partial}{\partial x^j}} \frac{\partial}{\partial x^k}
\end{aligned} \tag{5.27}$$

The last term contains a vector field on M which can be expanded in the basis $\partial/\partial x^i$, i.e.

$$\nabla_{\frac{\partial}{\partial x^j}} \frac{\partial}{\partial x^k} = \sum_i^n \Gamma^i_{jk} \frac{\partial}{\partial x^i} \tag{5.28}$$

where Γ^i_{jk} are called the *connection coefficients*. Not surprisingly we can make a special choice such that these coefficients are actually equal to the Christoffel symbols, but at the moment we will not make this assumption. In terms of these coefficients we can now write

$$\begin{aligned}
\nabla_v w &= \sum_{j,k}^n v^j \frac{\partial w^k}{\partial x^j} \frac{\partial}{\partial x^k} + \sum_{i,j,k}^n w^k v^j \Gamma^i_{jk} \frac{\partial}{\partial x^i} = \sum_{j,k} v^j \left(\frac{\partial w^k}{\partial x^j} + \sum_i^n \Gamma^k_{ji} v^i \right) \frac{\partial}{\partial x^k} \\
&= \sum_k (\nabla_v w)^k \frac{\partial}{\partial x^k}
\end{aligned} \tag{5.29}$$

in which we recovered the coefficients of Eq.(5.22). However, we have not assumed that we had a metric so at this moment we do not need to assume that the connection coefficients are symmetric, i.e. $\Gamma^k_{ij} = \Gamma^k_{ji}$. They do, however, transform exactly as the Christoffel symbols in Eq.(4.19). Let us check this. First we introduce the short notation

$$\nabla_{\frac{\partial}{\partial x^i}} = \nabla_i$$

then Eq.(5.28) is written as

$$\nabla_i \frac{\partial}{\partial x^j} = \sum_k^n \Gamma^k_{ij} \frac{\partial}{\partial x^k} \tag{5.30}$$

Similarly in a coordinate system y^j we have

$$\nabla'_\alpha \frac{\partial}{\partial y^\beta} = \sum_\gamma^n \Gamma'^{\gamma}_{\alpha\beta} \frac{\partial}{\partial y^\gamma} \tag{5.31}$$

Then

$$\nabla'_\alpha \frac{\partial}{\partial y^\beta} = \nabla'_\alpha \left(\sum_k^n \frac{\partial x^k}{\partial y^\beta} \frac{\partial}{\partial x^k} \right) = \sum_k^n \left(\frac{\partial}{\partial y^\alpha} \left(\frac{\partial x^k}{\partial y^\beta} \right) + \frac{\partial x^k}{\partial y^\beta} \nabla'_\alpha \frac{\partial}{\partial x^k} \right) \tag{5.32}$$

Now since

$$\nabla'_\alpha = \nabla_{\frac{\partial}{\partial y^\alpha}} = \nabla_{\sum_l \frac{\partial x^l}{\partial y^\alpha} \frac{\partial}{\partial x^l}} = \sum_l^n \frac{\partial x^l}{\partial y^\alpha} \nabla_l$$

it follows from Eq.(5.32) that

$$\begin{aligned}
\nabla'_\alpha \frac{\partial}{\partial y^\beta} &= \sum_k^n \frac{\partial^2 x^k}{\partial y^\alpha \partial y^\beta} \frac{\partial}{\partial x^k} + \sum_{k,l}^n \frac{\partial x^k}{\partial y^\beta} \frac{\partial x^l}{\partial y^\alpha} \nabla_l \frac{\partial}{\partial x^k} = \sum_k^n \frac{\partial^2 x^k}{\partial y^\alpha \partial y^\beta} \frac{\partial}{\partial x^k} + \sum_{k,l,m}^n \frac{\partial x^k}{\partial y^\beta} \frac{\partial x^l}{\partial y^\alpha} \Gamma^m_{lk} \frac{\partial}{\partial x^m} \\
&= \sum_{k,\gamma}^n \frac{\partial^2 x^k}{\partial y^\alpha \partial y^\beta} \frac{\partial y^\gamma}{\partial x^k} \frac{\partial}{\partial y^\gamma} + \sum_{k,l,m,\gamma}^n \Gamma^m_{lk} \frac{\partial x^k}{\partial y^\beta} \frac{\partial x^l}{\partial y^\alpha} \frac{\partial y^\gamma}{\partial x^m} \frac{\partial}{\partial y^\gamma}
\end{aligned} \tag{5.33}$$

Comparison of this expression with Eq.(5.31) then yields

$$\Gamma'^{\gamma}_{\alpha\beta} = \sum_{k,l,m}^n \Gamma^m_{lk} \frac{\partial x^l}{\partial y^\alpha} \frac{\partial x^k}{\partial y^\beta} \frac{\partial y^\gamma}{\partial x^m} + \sum_k^n \frac{\partial^2 x^k}{\partial y^\alpha \partial y^\beta} \frac{\partial y^\gamma}{\partial x^k} \tag{5.34}$$

which is identical to the transformation law for the Christoffel symbols of Eq.(4.19). However, we deduced it without making any reference to a metric.
We can further look at the connection $\nabla_v w$ in a different way. We have seen in Eq.(5.19) that the coefficients $w^k_{;j}$ transform as a mixed tensor of type $(1,1)$. This is not so surprising since $\nabla_v w$ is a vector and can therefore act on a covector . Moreover it is linear in v (see Eq.(5.25)). Therefore, for a given vector field w we can define a tensor $\nabla w \in \mathcal{T}^1_1(T_pM)$ by

$$\nabla w\,(v,u) = \nabla_v w\,(u) \tag{5.35}$$

where v is a vector field and u a covector field. In components this means the following. If

$$v = \sum_j^n v^j \frac{\partial}{\partial x^j} \quad , \quad u = \sum_j^n u_j dx^j$$

then

$$\begin{aligned}
\nabla w(v,u) &= \sum_j^n (\nabla_v w)^j \frac{\partial}{\partial x^j}(u) = \sum_j^n (\nabla_v w)^j u_j = \sum_{j,k}^n \left(\frac{\partial w^j}{\partial x^k} + \sum_i^n w^i \Gamma^j_{ki} \right) v^k u_j \\
&= \sum_{j,k}^n w^j_{;k}\, v^k u_j = \sum_{j,k}^n w^j_{;k}\, dx^k \otimes \frac{\partial}{\partial x^j}(v,u)
\end{aligned}$$

and therefore

$$\nabla w = \sum_{j,k}^{n} w^j_{;k}\, dx^k \otimes \frac{\partial}{\partial x^j} \tag{5.36}$$

The operator ∇ is usually called the connection or covariant derivative operator. Here we defined the covariant derivative on a vector field. We can, however, extend the definition of ∇ and ∇_v such that they can act on arbitrary tensor fields.
Let v be a vector field and let ∇ be a connection on vector fields. Then there is a unique operator $A \to \nabla_v A$ from tensor fields to tensor fields, preserving the type (k, l) such that

$$\nabla_v f = v(f) \tag{5.37}$$

$\nabla_v w$ is the vector field given by the connection ∇ (5.38)

$$\nabla_v(\lambda A) = \lambda \nabla A \quad \lambda \in \mathbb{R} \tag{5.39}$$

$$\nabla_v(A \otimes B) = \nabla_v A \otimes B + A \otimes \nabla_v B \tag{5.40}$$

For any contraction C we have $\nabla_v \circ C = C \circ \nabla_v$ (5.41)

Here f is a function and for any tensor A and a function f we define $f \otimes A = fA$. With these conditions we can construct the required mapping. Let us first consider the vector field w and covector field u and construct the tensor

$$A = u \otimes w = \sum_{j,k}^{n} u_j w^k dx^j \otimes \frac{\partial}{\partial x^k} = \sum_{j,k}^{n} A^k_j\, dx^j \otimes \frac{\partial}{\partial x^k}$$

The contraction of A is the scalar

$$\sum_{k}^{n} A^k_k = \sum_{k}^{n} u_k\, w^k$$

Condition (5.41) tells us that ∇_v should commute with tensor contractions C. First of all we have using Eq.(5.37) that

$$\nabla_v \circ C(A) = \nabla_v(\sum_{k}^{n} u_k\, w^k) = \sum_{k}^{n} v(u_k w^k) = \sum_{j,k}^{n} v^j \frac{\partial}{\partial x^j}(u^k w^k) \tag{5.42}$$

Next we will evaluate $C \circ \nabla_v$. We have using condition (5.40) that

$$\nabla_v A = \nabla_v(u \otimes w) = \nabla_v\, u \otimes w + u \otimes \nabla_v\, w \tag{5.43}$$

From condition (5.38) we have that

$$\nabla_v\, w = \sum_{j,k}^{n} w^k_{;j} v^j \frac{\partial}{\partial x^k}$$

We do not yet know the explicit form of $\nabla_v u$ but we know it will be again a covector field and we can therefore write

$$\nabla_v u = \sum_{l}^{n} (\nabla_v u)_l\, dx^l \tag{5.44}$$

Inserting this into Eq.(5.43) then gives

$$\nabla_v A = \sum_{k,l}^{n} \left[(\nabla_v u)_l w^k + u_l (\nabla_v w)^k\right]\, dx^l \otimes \frac{\partial}{\partial x^k} = \sum_{k,l}^{n} (\nabla_v A)^k_l\, dx^l \otimes \frac{\partial}{\partial x^k}$$

The contraction of A therefore gives the scalar function.

$$
\begin{aligned}
C \circ \nabla_v A &= \sum_k^n (\nabla A)_k^k = \sum_k^n \left[(\nabla_v u)_k w^k + u_k (\nabla_v w)^k \right] = \sum_k^n (\nabla_v u)_k w^k + \sum_{j,k}^n u_k w^k_{;j} v^j \\
&= \sum_k^n (\nabla_v u)_k w^k + \sum_{j,k}^n u_k \left(\frac{\partial w^k}{\partial x^j} + \sum_l w^l \Gamma^k_{jl} \right) v^j \\
&= \sum_{j,k}^n v^j \frac{\partial}{\partial x^j}(u_k w^k) - \sum_{j,k}^n w^k \frac{\partial u_k}{\partial x^j} v^j + \sum_k^n (\nabla_v u)_k w^k + \sum_{j,k,l} u_k w^l \Gamma^k_{jl} v^j
\end{aligned}
$$

This expression must be equal to Eq.(5.42) and by comparison we therefore find that

$$
0 = \sum_k^n \left[(\nabla_v u)_k - \sum_j^n v^j \left(\frac{\partial u_k}{\partial x^j} - \sum_l^n u_l \Gamma^l_{jk} \right) \right] w^k
$$

Since this must be valid for any set of coefficients w^k we find that

$$
(\nabla_v u)_k = \sum_j^n u_{k;j} v^j \tag{5.45}
$$

where we defined

$$
u_{k;j} = \frac{\partial u_k}{\partial x^j} - \sum_l^n u_l \Gamma^l_{jk} \tag{5.46}
$$

In particular if $v = \partial / \partial x^j$ we have from Eq.(5.44) that

$$
\nabla_j u = \sum_k^n w_{k;j} dx^k = \sum_k^n \left(\frac{\partial u_k}{\partial x^j} - \sum_l^n u_l \Gamma^l_{jk} \right) dx^k \tag{5.47}
$$

If we further specialize to $u = dx^i$ we see from this equation that

$$
\nabla_j dx^i = - \sum_k^n \Gamma^i_{jk} \, dx^k \tag{5.48}
$$

Together with Eq.(5.30) this can be used to calculate the derivative of a tensor of general type (p, q). If

$$
A = \sum_{\substack{i_1 \dots i_p \\ j_1 \dots j_q}}^n A^{j_1 \dots j_q}_{i_1 \dots i_p} \, dx^{i_1} \otimes \dots \otimes dx^{i_p} \otimes \frac{\partial}{\partial x^{j_1}} \otimes \dots \otimes \frac{\partial}{\partial x^{j_q}}
$$

then as a consequence of conditions (5.37) and (5.40) we have for $v = \partial/\partial x^k$ that

$$\begin{aligned}
\nabla_k A = & \sum_{i_1\dots i_p, j_1\dots j_q}^{n} \frac{\partial A^{j_1\dots j_q}_{i_1\dots i_p}}{\partial x^k}\, dx^{i_1}\otimes\dots\otimes dx^{i_p}\otimes\frac{\partial}{\partial x^{j_1}}\otimes\dots\otimes\frac{\partial}{\partial x^{j_q}} \\
& + \sum_{i_1\dots i_p, j_1\dots j_q}^{n} A^{j_1\dots j_q}_{i_1\dots i_p}\Big[\nabla_k dx^{i_1}\otimes\dots\otimes dx^{i_p}\otimes\frac{\partial}{\partial x^{j_1}}\otimes\dots\otimes\frac{\partial}{\partial x^{j_q}} \\
& \quad + dx^{i_1}\otimes\nabla_k dx^{i_2}\otimes\dots\otimes dx^{i_p}\otimes\frac{\partial}{\partial x^{j_1}}\otimes\dots\otimes\frac{\partial}{\partial x^{j_q}} + \dots \\
& \quad + dx^{i_1}\otimes\dots\otimes dx^{i_p}\otimes\nabla_k\frac{\partial}{\partial x^{j_1}}\otimes\dots\otimes\frac{\partial}{\partial x^{j_q}} + \dots \\
& \quad + dx^{i_1}\otimes\dots\otimes dx^{i_p}\otimes\frac{\partial}{\partial x^{j_1}}\otimes\dots\otimes\nabla_k\frac{\partial}{\partial x^{j_q}}\Big] \\
= & \sum_{i_1\dots i_p, j_1\dots j_q}^{n} A^{j_1\dots j_q}_{i_1\dots i_p;k}\, dx^{i_1}\otimes\dots\otimes dx^{i_p}\otimes\frac{\partial}{\partial x^{j_1}}\otimes\dots\otimes\frac{\partial}{\partial x^{j_q}}
\end{aligned}$$

where we used Eqs.(5.30) and (5.48) and relabeled some indices. We further defined

$$\begin{aligned}
A^{j_1\dots j_q}_{i_1\dots i_p;k} = & \frac{\partial A^{j_1\dots j_q}_{i_1\dots i_p}}{\partial x^k} - \sum_l^n A^{j_1\dots j_q}_{l i_2\dots i_p}\Gamma^l_{k i_1} - \dots - \sum_l^n A^{j_1\dots j_q}_{i_1\dots i_{p-1} l}\Gamma^l_{k i_p} \\
& + \sum_l^n A^{l j_2\dots j_q}_{i_1\dots i_p}\Gamma^{j_1}_{kl} + \dots + \sum_l^n A^{j_1\dots j_{q-1} l}_{i_1\dots i_p}\Gamma^{j_q}_{kl}
\end{aligned} \tag{5.49}$$

If you have some patience and a lot of paper then you can check that these coefficients transform properly as tensor coefficients under coordinate transformations. For the general operator $\nabla_v A$ we have

$$\nabla_v A = \sum_k^n v^k \nabla_k A = \sum_{\substack{k, i_1\dots i_p \\ j_1\dots j_q}}^{n} A^{j_1\dots j_q}_{i_1\dots i_p;k}\, v^k\, dx^{i_1}\otimes\dots\otimes dx^{i_p}\otimes\frac{\partial}{\partial x^{j_1}}\otimes\dots\otimes\frac{\partial}{\partial x^{j_q}} \tag{5.50}$$

Let us finally check that Eq.(5.49) indeed satisfies condition (5.41). If we, for example, contract the tensor A with respect to the indices i_1 and j_1 we obtain a new tensor B with components

$$B^{j_2\dots j_q}_{i_2\dots i_p} = \sum_{i_1}^n A^{i_1 j_2\dots j_q}_{i_1 i_2\dots i_p}$$

We need to verify the identity

$$B^{j_2\dots j_q}_{i_2\dots i_p;k} = \sum_{i_1}^n A^{i_1 j_2\dots j_q}_{i_1 i_2\dots i_p;k} \tag{5.51}$$

This follows from a direct computation. We have

$$\begin{aligned}
\sum_{i_1}^n A^{i_1 j_2\dots j_q}_{i_1 i_2\dots i_p;k} = & \frac{\partial}{\partial x^k}\left(\sum_{i_1}^n A^{i_1 j_2\dots j_q}_{i_1 i_2\dots i_p}\right) - \sum_{i_1,l}^n A^{i_1 j_2\dots j_q}_{l i_2\dots i_p}\Gamma^l_{k i_1} - \dots - \sum_{i_1,l}^n A^{i_1 j_2\dots j_q}_{i_1 i_2\dots i_{p-1} l}\Gamma^l_{k i_p} \\
& + \sum_{i_1,l}^n A^{l j_2\dots j_q}_{i_1 i_2\dots i_p}\Gamma^{i_1}_{kl} + \dots + \sum_{i_1,l}^n A^{i_1 j_2\dots j_{q-1} l}_{i_1 i_2\dots i_p}\Gamma^{j_q}_{kl}
\end{aligned} \tag{5.52}$$

We see that the second and the fourth term after the equality sign cancel. A quick inspection shows that the remaining terms sum up to $B^{j_2\ldots j_q}_{i_2\ldots i_p;k}$ such that covariant differentiation indeed commutes with contractions.
Let us finally give the analog of Eq.(5.35). Let $v_1,\ldots,v_p$ be vector fields and $u_1,\ldots,u_q$ be covector fields. Then we define

$$\nabla A(v,v_1,\ldots,v_p,u_1,\ldots,u_q) = \nabla_v\, A(v_1,\ldots,v_p,u_1,\ldots,u_q) \tag{5.53}$$

which implies that

$$\nabla A = \sum_{\substack{k,i_1\ldots i_p\\ j_1\ldots j_q}}^{n} A^{j_1\ldots j_q}_{i_1\ldots i_p;k}\; dx^k \otimes dx^{i_1} \otimes \ldots \otimes dx^{i_p} \otimes \frac{\partial}{\partial x^{j_1}} \otimes \ldots \otimes \frac{\partial}{\partial x^{j_q}} \tag{5.54}$$

So ∇ is a mapping from tensor fields of type (p,q) to tensor fields of type $(p+1,q)$.

5.3 Tensors along curves and particle motion revisited

So far we kept the connection coefficients general. The familiar Christoffel symbols that we used in the description of particle motion are recovered when we require that $\Gamma^k_{ij} = \Gamma^k_{ji}$ and that the covariant derivative of the metric tensor vanishes, i.e.

$$\nabla g = \sum_{i,j,k}^{n} g_{ij;k}\; dx^k \otimes dx^i \otimes dx^j = 0 \tag{5.55}$$

This condition yields explicitly using (5.49)

$$0 = g_{ij;k} = \frac{\partial g_{ij}}{\partial x^k} - \sum_l^n g_{lj}\Gamma^l_{ki} - \sum_l^n g_{il}\Gamma^l_{kj}$$

This yields

$$\frac{\partial g_{ij}}{\partial x^k} = [ik,j] + [jk,i] \tag{5.56}$$

where we defined

$$[ik,j] = \sum_l^n g_{lj}\,\Gamma^l_{ik}$$

and used the symmetry requirement $\Gamma^k_{ij} = \Gamma^k_{ji}$. By relabeling we then find

$$[ij,k] = \frac{1}{2}\left(\frac{\partial g_{ik}}{\partial x^j} + \frac{\partial g_{jk}}{\partial x^i} - \frac{\partial g_{ij}}{\partial x^k}\right)$$

and

$$\Gamma^k_{ij} = \sum_l^n g^{kl}[ij,l]$$

which is the familiar expression for the Christoffel symbols of the second kind. This connection is called the Riemann connection and arises from the condition (5.55). We will look into the geometrical meaning of this condition below. Let us, however, first deduce some consequences.

These follow from a combination of the conditions (5.40) and (5.41). For example, if A is a tensor of type $(2,2)$ and B a tensor of type $(1,1)$ then we have

$$\left(A^{j_1 j_2}_{i_1 i_1} B^{j_3}_{i_3}\right)_{;k} = A^{j_1 j_2}_{i_1 i_1;k} B^{j_3}_{i_3} + A^{j_1 j_2}_{i_1 i_1} B^{j_3}_{i_3;k}$$
$$\left(\sum_l^n A^{j_1 l}_{l i_1}\right)_{;k} = \sum_l^n A^{j_1 l}_{l i_1;k}$$

In particular, if we take the contraction of the metric tensor g with a vector v we have

$$\left(\sum_j^n g_{ij} v^j\right)_{;k} = \sum_j^n \left(g_{ij} v^j\right)_{;k} = \sum_j^n \left(g_{ij;k} v^j + g_{ij} v^j_{;k}\right) = \sum_j^n g_{ij} v^j_{;k} \tag{5.57}$$

as a consequence of the condition $g_{ij;k} = 0$. Therefore lowering the index of a vector commutes with taking the covariant derivative. The same applies when we want to raise indices. To see this we consider the identity

$$\delta^i_j = \sum_l^n g^{il} g_{lj} \tag{5.58}$$

Now for the tensor

$$\delta = \sum_{i,j}^n \delta^i_j \, dx^j \otimes \frac{\partial}{\partial x^i}$$

we have according to Eq.(5.49)

$$\delta^i_{j;k} = \frac{\partial}{\partial x^k} \delta^i_j - \sum_l^n \delta^i_l \Gamma^l_{kj} + \sum_l^n \delta^l_j \Gamma^i_{kl} = -\Gamma^i_{kj} + \Gamma^i_{kj} = 0$$

and hence from Eq.(5.58) we have

$$0 = \delta^i_{j;k} = \sum_l^n \left(g^{il}_{;k} g_{lj} + g^{il} g_{lj;k}\right) = \sum_l^n g^{il}_{;k} g_{lj}$$

Multiplying by g^{jm} and summing over j then yields $g^{im}_{;k} = 0$. Therefore completely analogously to Eq.(5.57) we have that if w is a covector then

$$\left(\sum_j^n g^{ij} w_j\right)_{;k} = \sum_j^n g^{ij} w_{j;k} \tag{5.59}$$

and therefore raising indices commutes with covariant differentiation. Let us finally explain the geometrical meaning of the condition $\nabla g = 0$. Let us go back to the operator $\nabla_v w$. In case v is the tangent vector to a curve $c(t) = (x^1(t), \ldots, x^n(t))$ we have

$$v(t) = \sum_j^n v^j(t) \frac{\partial}{\partial x^j}$$

We introduce the notation

$$\frac{Dw}{dt} = \nabla_{\dot{c}}\, w \tag{5.60}$$

This derivative is precisely the one we considered in the introduction of this Chapter. As we explained before, if w is the tangent vector field to a surface then Dw/dt represents the change in the vector field w in the tangent plane as we move along the curve $c(t)$. In case $Dw/dt = 0$ then, as a consequence of definition (5.10), there is only a change normal to the surface. In that case, if x^k are surface coordinates, we have

$$0 = \frac{Dw}{dt} = \nabla_{\dot{c}}\, w = \sum_k^n (\nabla_{\dot{c}}\, w)^k \frac{\partial}{\partial x^k}$$

and hence

$$0 = (\nabla_{\dot{c}}\, w)^k = \sum_j^n w^k_{;j}\, \dot{x}^j = \sum_j^n \left(\frac{\partial w^k}{\partial x^j} + \sum_i^n w^i \Gamma^k_{ji} \right) \dot{x}^j \tag{5.61}$$

Let us consider a special case. When the surface is a plane in $\mathbb{R}^3$ then the induced metric on the plane is flat and $\Gamma^k_{ij} = 0$. In that case Eq.(5.61) yields

$$0 = \sum_j^n \frac{\partial w^k}{\partial x^j}\, \dot{x}^j = \frac{\partial}{\partial t} w^k(x(t))$$

and we find that $w^k(x(t))$ is constant. In that case $Dw/dt = 0$ implies that the vector field along the curve consists of parallel vectors as displayed in Fig.5.5

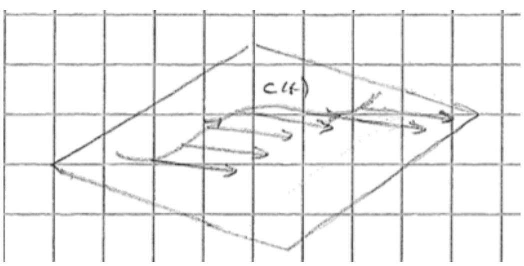

Figure 5.4: A parallel vector field along a path $c(t)$.

There is no change of the vectors in the plane and, in this particular case, the change normal to the surface is zero as well. In general we will call a vector field with the property $Dw/dt = 0$ a *parallel* vector field along the curve $c(t)$. Let us now go back to a generally curved surface and let u and w be parallel vector fields along $c(t)$, then

$$\begin{aligned}\frac{\partial}{\partial t} g(u,w) &= \frac{\partial}{\partial t} \sum_{i,j}^n g_{ij} u^i w^j = \sum_{i,j,k}^n \dot{x}^k \frac{\partial}{\partial x^k}(g_{ij} u^i w^j) = \sum_{i,j,k}^n \dot{x}^k (g_{ij} u^i w^j)_{;k} \\ &= \sum_{i,j,k}^n \dot{x}^k \left(g_{ij;k} u^i v^j + g_{ij} u^i_{;k} w^j + g_{ij} u^i w^j_{;k} \right) \\ &= \sum_{i,j}^n \left(g_{ij} \left(\frac{Du}{dt}\right)^i w^j + g_{ij} u^i \left(\frac{Dw}{dt}\right)^j \right) = 0 \end{aligned} \tag{5.62}$$

as a consequence of $g_{ij;k} = 0$ and the fact that we chose parallel vector fields. Therefore, if a connection is chosen to satisfy $\nabla g = 0$ then

$$g(u,w) = \text{constant}$$

for parallel vector fields u and w. In particular $g(u,u)$ is constant and therefore a Riemannian connection preserves the length of parallel vectors, as well as the angle between them.
So far we defined Dw/dt for a vector field w which was defined in a neighborhood of a curve $c(t)$. Suppose now that $w(t)$ is only defined on the curve $c(t)$ as is the case when $w(t) = \dot{c}(t)$ is the tangent vector to the curve $c(t)$. In such a case it does not make sense to write

$$\frac{\partial w^k}{\partial t} = \sum_j^n \frac{\partial w^k}{\partial x^j}\dot{x}^j$$

since $w^k(t)$ depends only on the values $x^j(t)$ on the curve an hence the partial derivatives $\partial w^k/\partial x^j$ are not known. For this case we therefore define

$$\frac{Dw}{dt} = \sum_k^n \left(\frac{Dw}{dt}\right)^k \frac{\partial}{\partial x^k} \tag{5.63}$$

where

$$\left(\frac{Dw}{dt}\right)^k = \frac{\partial w^k}{\partial t} + \sum_{i,j}^n w^i \Gamma^k_{ji}\dot{x}^j \tag{5.64}$$

Remember from our definitions that vectors are defined in single points, not necessarily as a vector field in the neighborhood of point. We only require that they act on functions f defined in the neighborhood of the point and that the coordinate system is defined in the same neighbourhood. More precisely if $x(t)$ is the coordinate of the point then

$$v(f)(x(t)) = \sum_j^n v^j(t)\frac{\partial f}{\partial x^j}(x(t))$$

If we introduce a new coordinate system $y(x)$ then the tangent vector to a curve $c(t)$ in the new coordinate system y is given by

$$v'^j(t) = \frac{\partial y^j}{\partial t} = \sum_k^n \frac{\partial y^j}{\partial x^k}\frac{\partial x^k}{\partial t} = \sum_k^n \frac{\partial y^j}{\partial x^k} v^k(t)$$

and the vector simply transforms as vector in the tangent space at $c(t)$. What we did for vectors also works for tensor fields that are only defined along a curve $c(t)$. For a tensor A of type (p,q) we define

$$\frac{DA}{dt} = \sum_{\substack{i_1\ldots i_p\\ j_1\ldots j_q}}^n \left(\frac{DA}{dt}\right)^{j_1\ldots j_q}_{i_1\ldots i_p}(t)\; dx^{i_1}\otimes\ldots\otimes dx^{i_p}\otimes\frac{\partial}{\partial x^{j_1}}\otimes\ldots\otimes\frac{\partial}{\partial x^{j_q}} \tag{5.65}$$

where

$$\begin{aligned}\left(\frac{DA}{dt}\right)^{j_1\ldots j_q}_{i_1\ldots i_p}(t) = \frac{\partial A^{j_1\ldots j_q}_{i_1\ldots i_p}}{\partial t}(t) + \sum_k^n \dot{x}^k \Bigg\{ &\sum_l^n A^{l j_2\ldots j_q}_{i_1\ldots i_p}\Gamma^{j_1}_{kl} + \ldots + \sum_l^n A^{j_1\ldots j_{q-1} l}_{i_1\ldots i_p}\Gamma^{j_q}_{kl} \\ &- \sum_l^n A^{j_1\ldots j_q}_{l i_2\ldots i_p}\Gamma^l_{k i_1} - \ldots - \sum_l^n A^{j_1\ldots j_q}_{i_1\ldots i_{p-1} l}\Gamma^l_{k i_p}\Bigg\}\end{aligned} \tag{5.66}$$

An important case of Eq.(5.64) is when $w = \dot{c}$ or equivalently $w^k(t) = \dot{x}^k$, which is the case in which we consider w to be tangent or velocity vector along the curve. In that case we obtain

$$\left(\frac{Dw}{dt}\right)^k = \ddot{x}^k + \sum_{i,j}^n \Gamma^k_{ij}\,\dot{x}^i\,\dot{x}^j \tag{5.67}$$

If we imagine the coordinates x^i to be surface coordinates (u^1, u^2) of a two-dimensional surface in $\mathbb{R}^3$ then we see from Eqs.(5.11) and (5.16) by taking w equal to the velocity vector $v = \dot{c}$ that Dv/dt represents the acceleration $a^{\|}$ of the particle parallel to the surface

$$a^{\|} = \frac{Dv}{dt} = \nabla_{\dot{c}}\, \dot{c} \tag{5.68}$$

Newton's law for the motion of a particle of mass m along the surface therefore becomes

$$m a^{\|} = m \frac{Dv}{dt} = F^{\|} \tag{5.69}$$

where $F^{\|}$ is the applied force. Let the force field be given by the gradient of a potential V in the surrounding three-dimensional space [1].

$$F = \sum_{j=1}^{3} F^j \frac{\partial}{\partial x^j} = -\sum_{j=1}^{3} \frac{\partial V}{\partial x^j} \frac{\partial}{\partial x^j}$$

For example, if the particle would be charged this could be an applied electric field. If mechanical forces keep the particle restricted to the surface we need to calculate the component $F^{\|}$ parallel to the surface. To do this we expand the force into the tangent vectors to the surface and the normal vector $\mathbf{n}$ as

$$F = F^{1,\|} \frac{\partial \mathbf{x}}{\partial u^1} + F^{2,\|} \frac{\partial \mathbf{x}}{\partial u^2} + F^{\perp}\, \mathbf{n}$$

We have

$$\langle \frac{\partial \mathbf{x}}{\partial u^i}, F \rangle = \sum_{j=1}^{2} \langle \frac{\partial \mathbf{x}}{\partial u^i}, \frac{\partial \mathbf{x}}{\partial u^j} \rangle F^{j,\|} = \sum_{i=1}^{2} g_{ij} F^{j,\|}$$

and we therefore find

$$F^{j,\|} = \sum_{k=1}^{2} g^{jk} \langle \frac{\partial \mathbf{x}}{\partial u^k}, F \rangle = -\sum_{k=1}^{2} g^{jk} \sum_{l=1}^{3} \frac{\partial x^l}{\partial u^k} \frac{\partial V}{\partial x^l} = -\sum_{k=1}^{2} g^{jk} \frac{\partial V}{\partial u_k} \tag{5.70}$$

Therefore we find that Newton's law (5.69) for the motion of the particle along the surface attains the form

$$m \left(\frac{Dv}{dt} \right)^j = m \left(\ddot{u}^j + \sum_{p,q=1}^{2} \Gamma^j_{pq}\, \dot{u}^p\, \dot{u}^q \right) = -\sum_{k=1}^{2} g^{jk} \frac{\partial V}{\partial u_k} \tag{5.71}$$

This is exactly Eq.(4.14) which we derived from the Lagrangian principle. Here we recovered the same equation by studying the forces and accelerations tangent to a surface. The equivalent equation (4.13) can be written in our new notation as

$$m \sum_{l=1}^{2} g_{kl} \left(\frac{Dv}{dt} \right)^l = -\frac{\partial V}{\partial x^k} \tag{5.72}$$

where both sides of the equation are now covectors rather than vectors. What we now want to show that we can also rewrite this equation as

$$m \left(\frac{Dv^\flat}{dt} \right)_k = -\frac{\partial V}{\partial x^k} \tag{5.73}$$

[1] There seems to be mismatch in the position of the indices. However, this is simply appearance because the metric is Euclidean $g_{ij} = \delta_{ij}$. We have $F^j = -\sum_k g^{jk} \partial V/\partial x^k = -\partial V/\partial x^j$.

where we take the covariant derivative of the covector $v^\flat$ obtained by lowering the indices on v. Since $v^\flat$ is a covector the covariant derivative should be taken according to Eq.(5.66) as

$$\left(\frac{Dv^\flat}{dt}\right)_k = \frac{\partial v_k}{\partial t} - \sum_{l,m=1}^{2} \dot{u}^m v_l \, \Gamma^l_{mk} \tag{5.74}$$

This is, in fact, an immediate consequence of Eq.(5.59) but let us check it explicitly for vectors and covectors defined along a curve. Rather tan restricting ourselves to two-dimensional surface we consider the case that we have n general coordinates. We have

$$\begin{aligned}\left(\frac{Dv^\flat}{dt}\right)_k &= \frac{\partial}{\partial t}\left(\sum_l^n g_{kl} v^l\right) - \sum_{l,m,p}^{n} \dot{u}^m g_{lp} v_p \, \Gamma^l_{mk} \\ &= \sum_{l,m}^{n} \frac{\partial g_{kl}}{\partial x^m} \dot{x}^m v^l + \sum_l^n g_{kl} \frac{\partial v^l}{\partial t} - \sum_{m,p}^{n} \dot{x}^m v^p [mk,p]\end{aligned} \tag{5.75}$$

where in the last term we used that (see Eq.(5.56))

$$\sum_l^n \Gamma^l_{mk} g_{lp} = \sum_{l,q}^{n} g^{lq} g_{lp} [mk,q] = \sum_q^n \delta^q_p [mk,q] = [mk,p]$$

If we further use that

$$\frac{\partial g_{kl}}{\partial x^m} = [km,l] + [lm,k]$$

then Eq.(5.75) becomes

$$\begin{aligned}\left(\frac{Dv^\flat}{dt}\right)_k &= \sum_l^n g_{kl} \frac{\partial v^l}{\partial t} + \sum_{l,m}^{n} ([km,l] + [lm,k] - [mk,l]) \dot{x}^m v^l \\ &= \sum_l^n g_{kl} \frac{\partial v^l}{\partial t} + \sum_{l,m}^{n} [lm,k] \dot{x}^m v^l = \sum_l^n g_{kl} \frac{\partial v^l}{\partial t} + \sum_{l,m,p}^{n} g_{kp} \Gamma^p_{lm} \dot{x}^m v^l \\ &= \sum_l^n g_{kl} \left(\frac{\partial v^l}{\partial t} + \sum_{i,j}^{n} \Gamma^l_{ij} \dot{x}^i v^j\right) = \sum_l^n g_{kl} \left(\frac{Dv}{dt}\right)^l\end{aligned} \tag{5.76}$$

which was exactly what we wanted to show. Therefore raising and lowering of indices commutes with D/dt.

5.4 Parallel vectors on a sphere

Let us finish with an explicit example that illustrates many of the things discussed. Let us parametrize a sphere with unit radius with the usual spherical coordinates

$$\begin{aligned} x^1 &= \cos\phi \sin\theta \\ x^2 &= \sin\phi \sin\theta \\ x^3 &= \cos\theta \end{aligned}$$

Then tangent vectors to the surface of the sphere (when imbedded in $\mathbb{R}^3$) are given by

$$\frac{\partial \mathbf{x}}{\partial \phi} = \begin{pmatrix} -\sin\phi\sin\theta \\ \cos\phi\sin\theta \\ 0 \end{pmatrix} \quad , \quad \frac{\partial \mathbf{x}}{\partial \theta} = \begin{pmatrix} \cos\phi\cos\theta \\ \sin\phi\cos\theta \\ -\sin\theta \end{pmatrix} \tag{5.77}$$

The metric tensor is then readily calculated from

$$g_{\phi\phi} = \langle \frac{\partial \mathbf{x}}{\partial \phi}, \frac{\partial \mathbf{x}}{\partial \phi} \rangle = \sin^2\theta \quad , \quad g_{\phi\theta} = \langle \frac{\partial \mathbf{x}}{\partial \phi}, \frac{\partial \mathbf{x}}{\partial \theta} \rangle = 0 \quad , \quad g_{\theta\theta} = \langle \frac{\partial \mathbf{x}}{\partial \theta}, \frac{\partial \mathbf{x}}{\partial \theta} \rangle = 1$$

and we therefore find that

$$g = d\theta \otimes d\theta + \sin^2\theta \, d\phi \otimes d\phi \tag{5.78}$$

Let us now calculate the Christoffel symbols of the first kind using Eq.(5.15). We have

$$\frac{\partial^2 \mathbf{x}}{\partial \phi^2} = \begin{pmatrix} -\cos\phi\sin\theta \\ -\sin\phi\sin\theta \\ 0 \end{pmatrix} \quad , \quad \frac{\partial^2 \mathbf{x}}{\partial \phi \partial \theta} = \begin{pmatrix} -\sin\phi\cos\theta \\ \cos\phi\cos\theta \\ 0 \end{pmatrix} \quad , \quad \frac{\partial^2 \mathbf{x}}{\partial \theta^2} = \begin{pmatrix} -\cos\phi\sin\theta \\ -\sin\phi\sin\theta \\ -\cos\theta \end{pmatrix}$$

and find that the only non-vanishing Christoffel symbols are given by

$$[\phi\phi, \theta] = -\sin\theta\cos\theta \quad , \quad [\phi\theta, \phi] = [\theta\phi, \phi] = \sin\theta\cos\theta \tag{5.79}$$

We can check some relations that we used, such as Eq.(5.56). We have, for instance, that

$$\begin{aligned} \frac{\partial g_{\phi\phi}}{\partial \theta} &= \frac{\partial}{\partial \theta}(\sin^2\theta) = 2\sin\theta\cos\theta = [\theta\phi, \phi] + [\phi\theta, \phi] \\ \frac{\partial g_{\phi\theta}}{\partial \phi} &= 0 = [\phi\phi, \theta] + [\theta\phi, \phi] \end{aligned}$$

The only non-vanishing Christoffel symbols of the second kind are given by

$$\begin{aligned} \Gamma^{\phi}_{\phi\theta} &= g^{\phi\phi}[\phi\theta, \phi] = \frac{1}{\sin^2\theta}\sin\theta\cos\theta = \cot\theta \\ \Gamma^{\theta}_{\phi\phi} &= g^{\theta\theta}[\phi\phi, \theta] = -\sin\theta\cos\theta \end{aligned}$$

This yields the following two equations for the geodesics

$$\begin{aligned} \ddot{\phi} + 2\,\dot{\theta}\dot{\phi}\,\cot\theta &= 0 \\ \ddot{\theta} - \dot{\phi}^2\,\sin\theta\cos\theta &= 0 \end{aligned}$$

From Eq.(5.17) we see that the equations for the covariant derivative of a vector field

$$v = v^\theta \frac{\partial}{\partial \theta} + v^\phi \frac{\partial}{\partial \phi}$$

are given by

$$\begin{aligned} (\nabla_\phi v) &= (\nabla_\phi v)^\theta \frac{\partial}{\partial \theta} + (\nabla_\phi v)^\phi \frac{\partial}{\partial \phi} \\ (\nabla_\theta v) &= (\nabla_\theta v)^\theta \frac{\partial}{\partial \theta} + (\nabla_\theta v)^\phi \frac{\partial}{\partial \phi} \end{aligned}$$

where

$$\begin{aligned} (\nabla_\phi v)^\theta &= v^\theta_{;\phi} = \frac{\partial v^\theta}{\partial \phi} + \Gamma^\theta_{\phi\phi} v^\phi = \frac{\partial v^\theta}{\partial \phi} - \sin\theta\cos\theta\, v^\phi \\ (\nabla_\phi v)^\phi &= v^\phi_{;\phi} = \frac{\partial v^\phi}{\partial \phi} + \Gamma^\phi_{\phi\theta} v^\theta = \frac{\partial v^\phi}{\partial \phi} + \cot\theta\, v^\theta \\ (\nabla_\theta v)^\theta &= v^\theta_{;\theta} = \frac{\partial v^\theta}{\partial \theta} \\ (\nabla_\theta v)^\phi &= v^\phi_{;\theta} = \frac{\partial v^\phi}{\partial \theta} + \Gamma^\phi_{\phi\theta} v^\phi = \frac{\partial v^\phi}{\partial \theta} + \cot\theta\, v^\phi \end{aligned}$$

We can check that we can properly lower the indices on these quantities. The covector $v^\flat$ obtained by lowering the indices on v is given by

$$v^\flat = v_\theta \, d\theta + v_\phi \, d\phi \tag{5.80}$$

where

$$\begin{aligned} v_\theta &= g_{\theta\theta} \, v^\theta = v^\theta \\ v_\phi &= g_{\phi\phi} \, v^\phi = \sin^2\theta \, v^\phi \end{aligned}$$

We can now calculate the covariant derivative of the covector using Eq.(5.46). We have

$$\begin{aligned} (\nabla_\phi v)_\theta &= v_{\phi;\theta} = \frac{\partial v_\phi}{\partial \theta} - \Gamma^\phi_{\phi\theta} v_\phi = \frac{\partial v_\phi}{\partial \theta} - \cot\theta \, v_\phi = \frac{\partial}{\partial\theta}(\sin^2\theta \, v^\phi) - \cot\theta \sin^2\theta \, v^\phi \\ &= \sin^2\theta \left[\frac{\partial v^\phi}{\partial \theta} + \cot\theta \, v^\phi \right] = g_{\phi\phi} \, v^\phi_{;\theta} \\ (\nabla_\phi v)_\phi &= v_{\phi;\phi} = \frac{\partial v_\phi}{\partial \phi} - \Gamma^\theta_{\phi\phi} v_\theta = \frac{\partial v_\phi}{\partial \phi} + \sin\theta\cos\theta \, v_\theta = \frac{\partial}{\partial\phi}\left(\sin^2\theta \, v^\phi\right) + \sin\theta\cos\theta \, v^\theta \\ &= \sin^2\theta \left[\frac{\partial v^\phi}{\partial \phi} + \cot\theta \, v^\theta \right] = g_{\phi\phi} \, v^\phi_{;\phi} \\ (\nabla_\theta v)_\theta &= \frac{\partial v_\theta}{\partial \theta} = \frac{\partial v^\theta}{\partial \theta} = g_{\theta\theta} \, v^\theta_{;\theta} \\ (\nabla_\theta v)_\phi &= v_{\theta;\phi} = \frac{\partial v_\theta}{\partial \phi} - \Gamma^\phi_{\theta\phi} v_\phi = \frac{\partial v_\theta}{\partial \phi} - \cot\theta \, v_\phi = \frac{\partial v_\theta}{\partial \phi} - \cot\theta \sin^2\theta \, v^\phi \\ &= \frac{\partial v_\theta}{\partial \phi} - \sin\theta\cos\theta \, v^\phi = g_{\theta\theta} \, v^\theta_{;\phi} \end{aligned}$$

Let us finally calculate the coefficients of Dv/dt. Using Eq.(5.64) we have

$$\left(\frac{Dv}{dt}\right)^\theta = \frac{\partial v^\theta}{\partial t} + \Gamma^\theta_{\phi\phi}\dot\phi \, v^\phi = \frac{\partial v^\theta}{\partial t} - \sin\theta\cos\theta \, \dot\phi \, v^\phi \tag{5.81}$$

$$\left(\frac{Dv}{dt}\right)^\phi = \frac{\partial v^\phi}{\partial t} + \Gamma^\phi_{\theta\phi}\dot\theta \, v^\phi + \Gamma^\phi_{\phi\theta}\dot\phi \, v^\theta = \frac{\partial v^\phi}{\partial t} + \cot\theta \, \dot\theta \, v^\phi + \cot\theta \, \dot\phi \, v^\theta \tag{5.82}$$

Let us from these two equations calculate the parallel vector fields satisfying $Dv/dt = 0$ in two cases. In the first case we take $\theta(t) = t$ and $\phi = \phi_0$ to be constant. Then Eqs.(5.81) and (5.82) yield

$$\begin{aligned} 0 &= \left(\frac{Dv}{dt}\right)^\theta = \frac{\partial v^\theta}{\partial t} \\ 0 &= \left(\frac{Dv}{dt}\right)^\phi = \frac{\partial v^\phi}{\partial t} + \cot(t) \, v^\phi \end{aligned}$$

The solution to these equations is

$$v^\theta = v_0 \quad , \quad v^\phi = \frac{K}{\sin t} \tag{5.83}$$

where v_0 and K are constants. We can check that the length of the vector is constant along the curve. We have

$$g(v,v) = g_{\theta\theta}(v^\theta)^2 + g_{\phi\phi}(v^\phi)^2 = v_0^2 + \sin^2 t \, \frac{K^2}{\sin^2 t} = v_0^2 + K^2$$

The reason that the component v^ϕ seems to grow as we move from the equator to the poles of the sphere is simply that are basis vector $\partial \mathbf{x}/\partial\phi$ has length $\sin\theta$ with decreases towards the pole. If we would introduce a normalized basis vector $\tilde{e}_\phi = (1/\sin\theta)\partial\mathbf{x}/\partial\phi$ then with respect to this basis we would have $\tilde{v}^\phi = K$. We therefore have the following picture for the vector field.

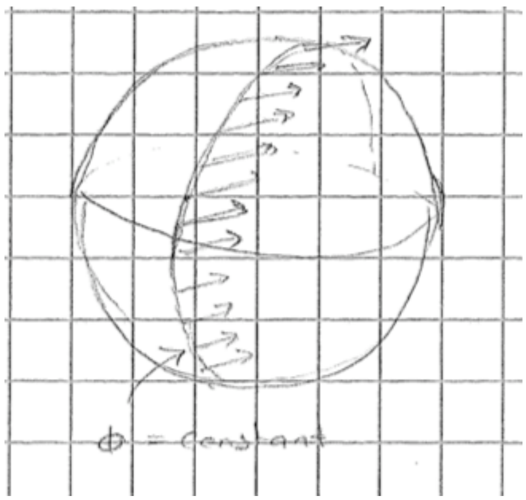

Figure 5.5: A parallel vector field along the curve $\phi = \phi_0$ and $\theta(t) = t$.

In the second case we take $\phi(t) = t$ and $\theta = \theta_0$ to be constant. In that case Eqs.(5.81) and (5.82) give

$$0 = \left(\frac{Dv}{dt}\right)^\theta = \frac{\partial v^\theta}{\partial t} - \sin\theta_0 \cos\theta_0 \, v^\phi$$
$$0 = \left(\frac{Dv}{dt}\right)^\phi = \frac{\partial v^\phi}{\partial t} + \cot\theta_0 \, v^\theta$$

From the first equation we find

$$v^\phi = \frac{1}{\sin\theta_0\cos\theta_0}\frac{\partial v^\theta}{\partial t}$$

which inserted into the second equation yields

$$\frac{\partial^2 v^\theta}{\partial t^2} = -\cos^2\theta_0 \, v^\theta$$

The general solution to this equation is

$$v^\theta(t) = A\cos(\omega t) + B\sin(\omega t)$$

with $\omega = \cos\theta_0$ (let us take the northern hemisphere such that $\theta_0 \in [0, \pi/2]$ then $\cos\theta_0 > 0$) and hence

$$\begin{aligned} v^\phi(t) &= \frac{1}{\sin\theta_0\cos\theta_0}\frac{\partial v^\theta}{\partial t} = \frac{\omega}{\sin\theta_0\cos\theta_0}(-A\sin(\omega t) + B\cos(\omega t)) \\ &= \frac{1}{\sin\theta_0}(-A\sin(\omega t) + B\cos(\omega t)) \end{aligned}$$

Then

$$\begin{pmatrix} v^\theta \\ v^\phi \end{pmatrix} = \frac{A}{\sin\theta_0}\begin{pmatrix} \cos(\omega t)\sin\theta_0 \\ -\sin(\omega t) \end{pmatrix} + \frac{B}{\sin\theta_0}\begin{pmatrix} \sin(\omega t)\sin\theta_0 \\ \cos(\omega t) \end{pmatrix}$$

Let us, for example, take $A = 1$ and $B = 0$. Then

$$\begin{pmatrix} v^\theta \\ v^\phi \end{pmatrix} = \frac{1}{\sin\theta_0}\begin{pmatrix} \cos(\omega t)\sin\theta_0 \\ -\sin(\omega t) \end{pmatrix} \tag{5.84}$$

Let us check that the length of the vectors is preserved if we move along the curve. We have

$$g(v,v) = g_{\theta\theta}(v^\theta)^2 + g_{\phi\phi}(v^\phi)^2 = \cos^2(\omega t) + \sin^2\theta_0 \frac{\sin^2(\omega t)}{\sin^2\theta_0} = 1$$

which is constant. As a three-dimensional vector field (i.e. embedded in the surrounding $\mathbb{R}^3$) this has the form

$$\begin{aligned} v(t) = v^\theta(t)\frac{\partial \mathbf{x}}{\partial\theta}(t) + v^\phi(t)\frac{\partial \mathbf{x}}{\partial\phi}(t) &= \cos(\omega t)\begin{pmatrix} \cos t\cos\theta_0 \\ \sin t\cos\theta_0 \\ -\sin\theta_0 \end{pmatrix} - \frac{\sin(\omega t)}{\sin\theta_0}\begin{pmatrix} -\sin t\sin\theta_0 \\ \cos t\sin\theta_0 \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} \omega\cos(\omega t)\cos t + \sin(\omega t)\sin t \\ \omega\cos(\omega t)\sin t - \sin(\omega t)\cos t \\ -\sin\theta_0\cos(\omega t) \end{pmatrix} \end{aligned} \tag{5.85}$$

If our calculation was right then $\partial v/\partial$ must only have a component normal to the surface. Differentiating Eq.(5.85) we find

$$\frac{\partial v}{\partial t} = \begin{pmatrix} (1-\omega^2)\sin(\omega t)\cos t \\ (1-\omega^2)\sin(\omega t)\sin t \\ \omega\sin\theta_0\sin(\omega t) \end{pmatrix} = \sin(\omega t)\begin{pmatrix} \sin^2\theta_0\cos t \\ \sin^2\theta_0\sin t \\ \sin\theta_0\cos\theta_0 \end{pmatrix} = \sin(\omega t)\sin\theta_0\,\mathbf{n}(t)$$

where

$$\mathbf{n}(t) = \begin{pmatrix} \cos t\sin\theta_0 \\ \sin t\sin\theta_0 \\ \cos\theta_0 \end{pmatrix}$$

is a unit vector normal to the surface of the sphere. The change in the vector field is therefore indeed only normal to the surface. From Eq.(5.85) we see that in order to achieve this the vector field must rotate in the tangent plane. We can rewrite Eq.(5.85) as

$$v(t) = \cos(\omega t)\mathbf{e}_\theta(t) - \sin(\omega t)\mathbf{e}_\phi(t) \tag{5.86}$$

where we defined

$$\mathbf{e}_\theta(t) = \frac{\partial \mathbf{x}}{\partial\theta}(t) \quad , \quad \mathbf{e}_\phi(t) = \frac{1}{\sin\theta_0}\frac{\partial \mathbf{x}}{\partial\phi}(t) \tag{5.87}$$

to be unit tangent vectors to the sphere. Pictorially this looks like

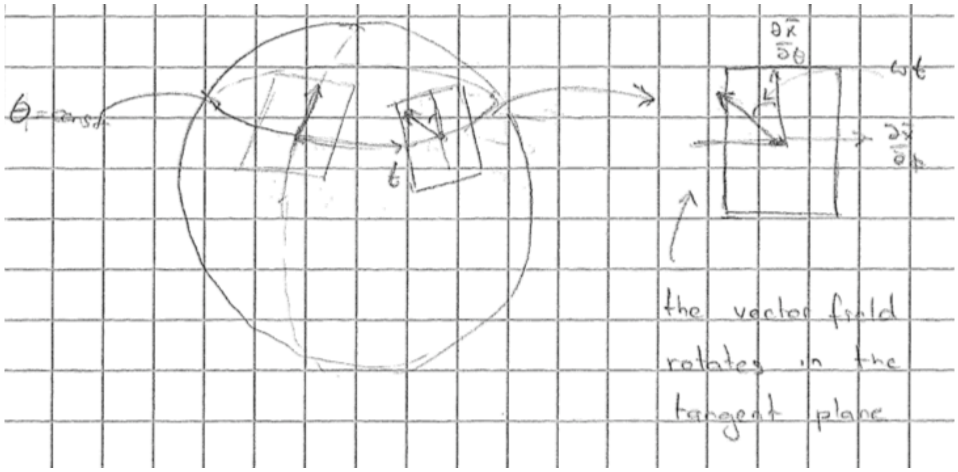

Figure 5.6: A parallel vector field along a the curve $\phi(t) = t$ and $\theta = \theta_0$.

Chapter 6

Relativistic mechanics and the Lorentz force

We present the relativistic force law, introduce the four-momentum and show how momentum and energy transform under Lorentz transformations. We show how the Lorentz force law as well as the electromagnetic field tensor F arise naturally from the assumption that the force is linear in the four-velocity. We derive the transformation laws of the electric and magnetic fields under a Lorentz transformation and work out two examples of particle motion in static fields.

6.1 Momentum and force in general coordinates

In Chapter 4 we discussed particle motion in general coordinates in Newtonian context. We found that the force law in general spatial coordinates has the form

$$F^k = m\left(\frac{Dv}{dt}\right)^k = m\,[\ddot{x}^k + \sum_{i,j=1}^{3} \Gamma^k_{ij}\,\dot{x}^i\,\dot{x}^j] \tag{6.1}$$

where

$$F^k = -\sum_{l=1}^{3} g^{kl}\frac{\partial V}{\partial x^l} \quad , \quad \dot{x}^k = \frac{\partial x^k}{\partial t} = v^k \tag{6.2}$$

The path of the particle $x(t)$ is a function of time which in a Newtonian theory is the same for any observer. In a Lorentz invariant theory space and time are transformed among each other. In this case only the proper time coordinate τ has an invariant meaning. It is therefore clear that the invariant generalization of Eq.(6.1) to general space-time transformations must be

$$F^k = m\left(\frac{Dv}{d\tau}\right)^k = m\,[\ddot{x}^k + \sum_{i,j=0}^{3} \Gamma^k_{ij}\,\dot{x}^i\,\dot{x}^j] \tag{6.3}$$

where

$$\dot{x}^k = \frac{\partial x^k}{\partial \tau} = v^k(\tau) \tag{6.4}$$

Eq.(6.3) looks completely identical to Eq.(6.1). We have only replaced the absolute time of the Newtonian theory by the proper time τ. We further deal with a four-dimensional space-time manifold rather than a three-dimensional spatial manifold. The vectors in Eq.(6.3) and (6.4) have four components and are therefore referred to a *four-vectors*. The simplest four-vector is the velocity four-vector $v^k(\tau)$ of Eq.6.4) which is just the tangent vector to the world line of the particle in space-time. In Chapter 1 we defined the world line in a Cartesian coordinate system but it is clear that we can do this in any coordinate system that we like. Let us, for the moment, stick to standard coordinates in a Lorentzian frame $x^k(\tau) = (x^0 = ct, x^1, x^2, x^3))$ then according to Eq.(1.64) we have

$$-c^2 d\tau^2 = -(dx^0)^2 + (dx^1)^2 + (dx^2)^2 + (dx^3)^2 \tag{6.5}$$

where dx^k are the infinitesimal differences between two space-time events that happen at the same spatial points in the reference frame that moves with the particle. From Eq.(6.5) it follows that

$$-c^2 = -\left(\frac{dx^0}{d\tau}\right)^2 + \left(\frac{dx^1}{d\tau}\right)^2 + \left(\frac{dx^2}{d\tau}\right)^2 + \left(\frac{dx^3}{d\tau}\right)^2 \tag{6.6}$$

If we introduce the metric tensor for the Minkowski metric

$$g = \sum_{i,j=0}^{3} g_{ij}\, dx^i \otimes dx^j \tag{6.7}$$

where

$$g_{ij} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

and write

$$v(\tau) = \sum_{j=0}^{3} v^j(\tau)\frac{\partial}{\partial x^j} = \sum_{j=0}^{3} \frac{\partial x^j}{\partial \tau}\frac{\partial}{\partial x^j} \tag{6.8}$$

then Eq.(6.5) is equivalent to

$$-c^2 = g(v, v) \tag{6.9}$$

From this equation we see that the tangent vector v to the world line has constant length. This means that the world line curve, up to a factor of c, is parametrized by arc length (see Eq.(4.34)). if we transform to arbitrary new coordinates (y^0, y^1, y^2, y^3) then Eq.(6.6) becomes

$$-c^2 = \sum_{i,j,k,l=0}^{3} g_{ij}\frac{\partial x^i}{\partial y^k}\frac{\partial x^j}{\partial y^l}\frac{\partial y^k}{\partial \tau}\frac{\partial y^l}{\partial \tau} = \sum_{k,l=0}^{3} g'_{kl}\frac{\partial y^k}{\partial \tau}\frac{\partial y^l}{\partial \tau}$$

where

$$g'_{kl} = \sum_{i,j=0}^{3} g_{ij}\frac{\partial x^i}{\partial y^k}\frac{\partial x^j}{\partial y^l}$$

For instance, when transforming the spatial Cartesian coordinates to spherical coordinates (x^0, r, θ, ϕ) we have

$$-c^2 = -\left(\frac{dx^0}{d\tau}\right)^2 + \left(\frac{dr}{d\tau}\right)^2 + r^2\left(\frac{d\theta}{d\tau}\right)^2 + r^2\sin^2\theta\left(\frac{d\phi}{d\tau}\right)^2$$

In the absence of gravitational forces (or more precisely for a vanishing Riemann tensor) the metric can always *globally* be transformed to the Minkowskian form. This metric defines a equivalence class of physically equivalent metrics up to a coordinate transformation (which for instance includes the metric of the rotating disc discussed in Chapter 2). In the presence of gravitational fields Minkowskian form can only be achieved *locally*, describing non-rotating and free falling frames. These are the local inertial frames described in Chapter 2. The global metric

is determined by the mass and energy distribution and represents again an equivalence class of metrics up to a coordinate transformation.
The next step is to give a description of the four-force in Eq.(6.3). For classical (i.e. non-quantum) applications the most important are electromagnetism and gravity. The simplest case is gravity since in that case we simply have $F^k = 0$ as particles in a gravity field move freely along geodesics in space-time. The only way gravity enters is via the Christoffel symbols Γ^k_{ij} which are determined by the metric g_{ij}. We see that gravity is actually no force. This is exactly as the constrained motion of particles along surfaces which we considered in Chapter 4. The deviation from geodesic motion is caused by true forces not related to the curvature of space the particles moves in. Consider, for example, the motion of two electrically charged particles on the surface of torus. The interaction between the particles will push them away from free geodesic motion. We will soon see what form F^k attains in the case of electromagnetism, but let us for the moment keep the vector F^k general.
We start by defining the *four-momentum* vector in arbitrary coordinates x^k

$$p^k = m\frac{\partial x^k}{\partial \tau} = mv^k(\tau) \tag{6.10}$$

where m is the mass (or more appropriately the rest mass of the particle). The *four-force* F^k is then defined to be

$$F^k = \left(\frac{Dp}{d\tau}\right)^k = m\left(\frac{Dv}{d\tau}\right)^k \tag{6.11}$$

i.e. it is the covariant derivative of the vector p along the path $x^k(\tau)$ of the particle. This is exactly Eq.(6.3). Our definition is not useful until we have also given a description of the physical origin of the force. Let us, however, start by deriving a condition on F^k. From Eq.(6.9) it follows that

$$\begin{aligned}
0 &= \frac{\partial}{\partial \tau} g(v,v) = \frac{\partial}{\partial \tau}\sum_{i,j} g_{ij}v^iv^j = \sum_{i,j,k}\frac{\partial g_{ij}}{\partial x^k}\frac{\partial x^k}{\partial \tau}v^iv^j + \sum_{i,j} g_{ij}\left(\frac{\partial v^i}{\partial t}v^j + v^i\frac{\partial v^j}{\partial \tau}\right)\\
&= \sum_{i,j,k}([ik,j]+[jk,i])\frac{\partial x^k}{\partial \tau}v^iv^j + \sum_{i,j} g_{ij}\left(\frac{\partial v^i}{\partial t}v^j + v^i\frac{\partial v^j}{\partial \tau}\right)\\
&= \sum_{i,j} g_{ij}\left(\frac{\partial v^i}{\partial t}v^j + v^i\frac{\partial v^j}{\partial \tau}\right) + \sum_{i,j,k,l}(\Gamma^l_{ik}g_{lj} + \Gamma^l_{jk}g_{li})\frac{\partial x^k}{\partial \tau}v^iv^j\\
&= \sum_{i,j} g_{ij}\left\{\left(\frac{\partial v^i}{\partial \tau} + \sum_{k,l}\Gamma^i_{lk}v^l\frac{\partial x^k}{\partial \tau}\right)v^j + v^i\left(\frac{\partial v^j}{\partial \tau} + \sum_{k,l}\Gamma^j_{lk}v^l\frac{\partial x^k}{\partial \tau}\right)\right\}\\
&= \sum_{i,j} g_{ij}\left\{\left(\frac{Dv}{d\tau}\right)^i v^j + v^i\left(\frac{Dv}{d\tau}\right)^j\right\} = 2\sum_{i,j} g_{ij}\left(\frac{Dv}{d\tau}\right)^i v^j
\end{aligned}$$

It therefore follows that

$$0 = g(v, \frac{Dv}{d\tau}) \tag{6.12}$$

and consequently from Eq.(6.11) that

$$0 = g(v,F) = \sum_{i,j=0}^{3} g_{ij}v^iF^j = \sum_{i=0}^{3} F_iv^i \tag{6.13}$$

We therefore see that the contraction of the covector $F^\flat$ with components

$$F_i = \sum_{j=0}^{3} g_{ij}F^j$$

with the vector v is zero. We can also write this as $F^\flat(v) = 0$. Since the momentum p and force F are vectors, it is clear how they transform under general coordinate transformations. However, to get some insight into the physical meaning of these vectors we will return to the case of standard Lorentz frames in the next Section.

6.2 Lorentz transformation of momentum and energy

Let us now again consider the case that we have Cartesian coordinates and let us describe the motion of a particle in terms of the time-variable t in some Lorentz frame. Then the world line is given by

$$x(t) = (x^0(t) = ct, x^1(t), x^2(t), x^3(t)) \tag{6.14}$$

It is clear that there is a one-to-one correspondence between t and the proper time τ since t grows monotonically with τ. We can therefore regard $t(\tau)$ as a function of τ. If we differentiate Eq.(6.14) with respect to τ we find

$$v^k(\tau) = \frac{\partial x^k}{\partial \tau} = \frac{\partial x^k}{\partial t}\frac{\partial t}{\partial \tau}$$

and therefore

$$v(\tau) = \left(c, \frac{\partial x^1}{\partial t}, \frac{\partial x^2}{\partial t}, \frac{\partial x^3}{\partial t}\right)\frac{\partial t}{\partial \tau} \tag{6.15}$$

The last three components within brackets are simply the components of the velocity as observed from the Lorentz frame in which the particle moves. From Eqs.(6.15) and (6.6) we then find

$$-c^2 = \left[-c^2 + \left(\frac{dx^1}{dt}\right)^2 + \left(\frac{dx^2}{dt}\right)^2 + \left(\frac{dx^3}{dt}\right)^2\right]\left(\frac{\partial t}{\partial \tau}\right)^2$$

and therefore

$$\frac{\partial t}{\partial \tau} = \frac{1}{\sqrt{1 - \frac{u(t)^2}{c^2}}} = \gamma \tag{6.16}$$

where we used that $\partial t/\partial \tau > 0$ and defined the velocity

$$u(t) = \sqrt{\left(\frac{dx^1}{dt}\right)^2 + \left(\frac{dx^2}{dt}\right)^2 + \left(\frac{dx^3}{dt}\right)^2}$$

From Eq.(6.15) we then see that

$$v(\tau) = \gamma(c, u^1, u^2, u^3) \tag{6.17}$$

where $u^i = \partial x^i/\partial t$ for $i = 1, 2, 3$. From Eq.(6.17) we then see that the momentum four-vector takes the form

$$p = mv = \gamma(mc, mu^1, mu^2, mu^3) \tag{6.18}$$

The three-vector

$$\mathbf{p}(t) = \gamma m\mathbf{u} = \frac{m\mathbf{u}}{\sqrt{1 - \frac{u^2}{c^2}}} \tag{6.19}$$

where $\mathbf{u} = (u^1, u^2, u^3)$ represents the generalization of the Newtonian momentum $\mathbf{p} = m\mathbf{u}$ to Lorentz frames. Let us see if we can attach some physical meaning to the component p^0 as well. If there is no force acting on the particle then $dp/d\tau = 0$ and $p(\tau)$ is a constant vector. For the three spatial components this implies conservation of three-momentum. We also know that the energy of a freely moving particle is conserved. It is therefore to be expected that this is described by the p^0 component in Eq.(6.18). From dimensional considerations we see that $p^0 = E/c$ where E is the energy. Eq.(6.18) then tells us that

$$E = p^0 c = \frac{mc^2}{\sqrt{1 - \frac{u^2}{c^2}}} \tag{6.20}$$

We can also arrive at this result in a different way. We will use the following familiar equation form classical mechanics

$$\frac{\partial E}{\partial t} = \mathbf{F} \cdot \mathbf{u} = m \frac{\partial \mathbf{u}}{\partial t} \cdot \mathbf{u} = \frac{\partial}{\partial t} \left(\frac{1}{2} m u^2 \right)$$

which says that the rate of change of the energy of a particle is equal to the work $\mathbf{F} \cdot \mathbf{u}$ done on the particle. Let us see what the equivalent of this equation is in special relativity. We consider Eq.(6.13) for the case of a Minkowskian metric such that

$$0 = F_0 v^0 + F_1 v^1 + F_2 v^2 + F_3 v^3 = -F^0 v^0 + F^1 v^1 + F^2 v^2 + F^3 v^3$$

From this equation we see that

$$\frac{\partial p^0}{\partial \tau} v^0 = F^0 v^0 = \sum_{j=1}^{3} \frac{\partial p^j}{\partial \tau} v^j \tag{6.21}$$

Further using

$$\frac{\partial p^k}{\partial \tau} = \frac{\partial p^k}{\partial t} \frac{\partial t}{\partial \tau} = \gamma \frac{\partial p^k}{\partial t}$$

on both sides of the equation, as well as Eq.(6.17), we see that

$$c \frac{\partial p^0}{\partial t} = \sum_{j=1}^{3} u^j \frac{\partial p^j}{\partial t} = \mathbf{u} \cdot \frac{\partial \mathbf{p}}{\partial t} = \mathbf{F} \cdot \mathbf{u} \tag{6.22}$$

We therefore find that

$$c \frac{\partial p^0}{\partial t} = \frac{\partial E}{\partial t}$$

and hence

$$E = p^0 c + K$$

where K is an integration constant. If we put $K = 0$ we recover Eq.(6.20). In case that $u \ll c$ we find from Eq.(6.20)

$$E = mc^2 + \frac{1}{2} m u^2 + \frac{3}{8} m \frac{u^4}{c^2} + \dots$$

The second term in this equation presents the familiar kinetic energy from Newtonian mechanics. The first term describes the energy content of a particle at zero velocity

$$E(u = 0) = mc^2$$

This is not just a constant since mass is not conserved in relativistic particle collisions, unlike in Newtonian mechanics. Mass is therefore proportional to energy content and not a separate physical quantity. In particular, we conclude that any type of energy is a source of gravity.
Since the four-momentum is a vector, i.e. more precisely

$$p(\tau) = \sum_{j=0}^{3} p^j(\tau) \frac{\partial}{\partial x^j} = \sum_{j=0}^{3} p^j(\tau) \frac{\partial}{\partial x^j} = \sum_{j,k=0}^{3} p^j(\tau) \frac{\partial y^k}{\partial x^j} \frac{\partial}{\partial y^k}$$

we have

$$p'^k(\tau) = \sum_{j=0}^{3} \frac{\partial y^k}{\partial x^j}\, p^j(\tau)$$

In case y^k are coordinates in a Lorentz frame O' moving with respect to our original frame O with constant speed v along the x^1 axis we have

$$\frac{\partial y^k}{\partial x^j} = \begin{pmatrix} \gamma & -\gamma\, v/c & 0 & 0 \\ -\gamma\, v/c & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \qquad \gamma = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}} \tag{6.23}$$

and therefore

$$p'^0 = \frac{p^0 - \frac{v}{c}p^1}{\sqrt{1-\frac{v^2}{c^2}}} \quad , \quad p'^1 = \frac{p^1 - \frac{v}{c}p^0}{\sqrt{1-\frac{v^2}{c^2}}} \tag{6.24}$$

and $p'^2 = p^2, p'^3 = p^3$. We can write this equivalently as

$$E' = \frac{E - vp^1}{\sqrt{1-\frac{v^2}{c^2}}} \quad , \quad p'^1 = \frac{p^1 - \frac{vE}{c^2}}{\sqrt{1-\frac{v^2}{c^2}}} \tag{6.25}$$

From Eq.(6.6) and the definition (6.10) we further obtain the useful relation

$$-m^2c^2 = -\left(\frac{E}{c}\right)^2 + \mathbf{p}^2$$

or

$$E = c\sqrt{m^2c^2 + \mathbf{p}^2} \tag{6.26}$$

where $\mathbf{p}$ is the three-momentum. This relation is often used in the study of particle collisions. For instance, when we have two incoming particles with four-momenta $(E_1/c, \mathbf{p}_1)$ and $(E_2/c, \mathbf{p}_2)$ and outgoing four-momenta $(E_3/c, \mathbf{p}_3)$ and $(E_4/c, \mathbf{p}_4)$ then the conservation of four-momentum tells us that

$$\sqrt{m_1^2c^2 + \mathbf{p}_1^2} + \sqrt{m_2^2c^2 + \mathbf{p}_2^2} = \sqrt{m_3^2c^2 + \mathbf{p}_3^2} + \sqrt{m_4^2c^2 + \mathbf{p}_4^2}$$
$$\mathbf{p}_1 + \mathbf{p}_2 = \mathbf{p}_3 + \mathbf{p}_4$$

Using these equations we could now study a large number of different collision processes. We will not do this here but instead continue to discuss the force law Eq.(6.3) in the electromagnetic case.

6.3 Lorentz force

To predict the motion of a particle under the influence of a force F^k we have to solve Eq.(6.3). However, we first need to describe how the force looks like at any space-time point. One of the simplest force laws that one could write down is that the change in four-velocity is simply proportional to the four-velocity. In terms of an equation

$$m\left(\frac{Dv}{dt}\right)^k = \tilde{q}\,\mathcal{F}^k(v) \tag{6.27}$$

where $\tilde{q}$ is a proportionality constant to be determined later and where $\mathcal{F}$ is a linear transformation that maps a four-vector to a four-vector, i.e.

$$\mathcal{F}(\alpha v_1 + \beta v_2) = \alpha\,\mathcal{F}(v_1) + \beta\,\mathcal{F}(v_2)$$

If we write out $\mathcal{F}$ in components we can write Eq.(6.27) as

$$m\left(\frac{Dv}{dt}\right)^k = \tilde{q}\sum_{l=0}^{3}\mathcal{F}^k_l\, v^l \tag{6.28}$$

We see that the coefficients $\mathcal{F}^k_l$ must be the components of a mixed tensor of type $(1,1)$, i.e.

$$\mathcal{F} = \sum_{k=0}^{3}\mathcal{F}^k_l\, dx^l \otimes \frac{\partial}{\partial x^k}$$

If we act with $\mathcal{F}$ on a vector v and a covector w we have

$$\mathcal{F}(v,w) = \sum_{k,l=0}^{3}\mathcal{F}^k_l v^l w_k = \sum_{k=0}^{3}(\mathcal{F}(v))^k w_k = (\mathcal{F}(v))(w)$$

where $\mathcal{F}(v)$ is the vector

$$\mathcal{F}(v) = \sum_{j=0}^{3}(\mathcal{F}(v))^k \frac{\partial}{\partial x^k} \quad , \quad (\mathcal{F}(v))^k = \sum_{l=0}^{3}\mathcal{F}^k_l v^l \tag{6.29}$$

Given $\mathcal{F}$ we can define the second order covariant tensor F acting on vectors u and v by

$$F(u,v) = g(u,\mathcal{F}(v)) \tag{6.30}$$

where g is the metric tensor. In components we have

$$F(u,v) = \sum_{i,j=0}^{3} g_{ij}u^i(\mathcal{F}(v))^j = \sum_{i,j,k=0}^{3} g_{ij}u^i\mathcal{F}^j_k v^k = \sum_{i,k=0}^{3} F_{ik}u^i v^k$$

where

$$F_{ik} = \sum_{j=0}^{3} g_{ij}\mathcal{F}^j_k \tag{6.31}$$

is obtained from $\mathcal{F}$ by lowering the upper index. The tensor F has the property

$$F(v,v) = g(v,\mathcal{F}(v)) = \frac{m}{\tilde{q}}\, g(v, \frac{Dv}{d\tau}) = 0 \tag{6.32}$$

where in the last step we used the property (6.12). If we assume that any (time-like) four-vector can be the four-velocity of a particle in field F then we have for four-vectors u and v that

$$F(u,v) + F(v,u) = \frac{1}{2}[F(u+v,u+v) - F(u-v,u-v)] = 0$$

and therefore

$$F(u,v) = -F(v,u)$$

It therefore follows that F is an anti-symmetric tensor, or in components

$$F_{ik} = -F_{ki} \tag{6.33}$$

Now using Eq.(6.31)

$$\sum_{l=0}^{3} g^{li} F_{ik} = \mathcal{F}^l_k$$

we can write Eq.(6.28) as

$$m\left(\frac{Dv}{dt}\right)^l = \tilde{q}\sum_{k=0}^{3}\mathcal{F}^l_k\, v^k = \tilde{q}\sum_{i,k=0}^{3} g^{li}F_{ik}\, v^k \tag{6.34}$$

or equivalently

$$m\left(\frac{Dv}{dt}\right)_i = \tilde{q}\sum_{k=0}^{3} F_{ik}\, v^k \tag{6.35}$$

Therefore from Eq.(6.27) we conclude that a linear relation between the four-velocity and its covariant derivative implies that the four-velocity and the covariant derivative of its covector are related by a rank two anti-symmetric covariant tensor. Since F_{ik} is anti-symmetric we can always write it in components as

$$F = \begin{pmatrix} 0 & -E_1 & -E_2 & -E_3 \\ E_1 & 0 & B_3 & -B_2 \\ E_2 & -B_3 & 0 & B_1 \\ E_3 & B_2 & -B_1 & 0 \end{pmatrix} \tag{6.36}$$

From the way that we derived the form of this expression we can not deduce the physical meaning of the six non-vanishing components $(E_1, E_2, E_3, B_1, B_2, B_3)$. We can just note that they may a different role since the coefficients (E_1, E_2, E_3) involve the mixed space-time components F_{0j} of the tensor whereas the coefficients (B_1, B_2, B_3) involve the purely spatial $F_{ij}, i, j = 1, 2, 3$ components of the tensor. The components (E_1, E_2, E_3) will be called the components of the electric field whereas the components (B_1, B_2, B_3) will be called the components of the magnetic field. With the form Eq.(6.36) we can write out Eq.(6.35) as

$$m\begin{pmatrix} \left(\frac{Dv}{d\tau}\right)_0 \\ \left(\frac{Dv}{d\tau}\right)_1 \\ \left(\frac{Dv}{d\tau}\right)_2 \\ \left(\frac{Dv}{d\tau}\right)_3 \end{pmatrix} = \tilde{q}\begin{pmatrix} -E_1v^1 - E_2v^2 - E_3v^3 \\ E_1v^0 + v^2B_3 - v^3B_2 \\ E_2v^0 - v^1B_3 + v^3B_1 \\ E_3v^0 + v^1B_2 - v^2B_1 \end{pmatrix}$$

and we therefore obtain the equations

$$m\left(\frac{Dv}{d\tau}\right)_0 = -\tilde{q}(E_1v^1 + E_2v^2 + E_3v^3) \tag{6.37}$$

$$m\begin{pmatrix} \left(\frac{Dv}{d\tau}\right)_1 \\ \left(\frac{Dv}{d\tau}\right)_2 \\ \left(\frac{Dv}{d\tau}\right)_3 \end{pmatrix} = \tilde{q}v^0\begin{pmatrix} E_1 \\ E_2 \\ E_3 \end{pmatrix} + \tilde{q}\begin{pmatrix} v^2B_3 - v^3B_2 \\ v^3B_1 - v^1B_3 \\ v^1B_2 - v^2B_1 \end{pmatrix} \tag{6.38}$$

To write these equations in a familiar form we take Cartesian coordinates in a Lorentz frame with Minkoskian metric. We then have

$$\begin{aligned}\left(\frac{Dv}{d\tau}\right)_i &= \sum_{j=0}^{3} g_{ij}\left(\frac{Dv}{d\tau}\right)^j = \sum_{j=0}^{3} g_{ij}\frac{\partial v^j}{\partial t}\frac{\partial t}{\partial \tau} = \sum_{j=0}^{3}\gamma\, g_{ij}\frac{\partial v^j}{\partial t} \\ &= \gamma(-\frac{\partial v^0}{\partial t}, \frac{\partial v^1}{\partial t}, \frac{\partial v^2}{\partial t}, \frac{\partial v^3}{\partial t}) \end{aligned} \tag{6.39}$$

We then have from Eq.(6.17) that $v(\tau) = \gamma(c, \mathbf{u})$ and therefore we can write Eqs.(6.37) and (6.38) as

$$m\frac{\partial}{\partial t}(c\,\gamma) = \tilde{q}(E_1 u^1 + E_2 u^2 + E_3 u^3)$$
$$m\frac{\partial}{\partial t}(\gamma\,\mathbf{u}) = \tilde{q}c\begin{pmatrix} E_1 \\ E_2 \\ E_3 \end{pmatrix} + \tilde{q}\begin{pmatrix} u^1 \\ u^2 \\ u^3 \end{pmatrix} \times \begin{pmatrix} B^1 \\ B^2 \\ B^3 \end{pmatrix}$$

If we now choose $\tilde{q} = q/c$ where q is the electric charge of the particle we recover the familiar equations

$$\frac{\partial E}{\partial t} = \frac{\partial}{\partial t}(\gamma\, mc^2) = q\,\mathbf{E}\cdot\mathbf{u} \tag{6.40}$$
$$\frac{\partial \mathbf{p}}{\partial t} = \frac{\partial}{\partial t}(\gamma\, m\mathbf{u}) = q(\mathbf{E} + \frac{1}{c}\,\mathbf{u}\times\mathbf{B}) \tag{6.41}$$

Eq.(6.41) is the famous Lorentz force law (we use Gaussian units) and Eq.(6.40) describes the change in energy of the particle when it is being accelerated or de-accelerated in the electric field $\mathbf{E}$ by a force $\mathbf{F} = q\mathbf{E}$.

6.4 Transformation of electric and magnetic fields

Another useful consequence of our derivation is that, since we know that F is a rank two covariant tensor, we can easily calculate ist components in a different coordinate frame and in this way determine the transformation law of the electric and magnetic fields under Lorentz transformations. We have

$$F = \sum_{k,l=0}^{3} F'_{kl}\, dy^k \otimes dy^l = \sum_{i,j=0}^{3} F_{ij}\, dx^i \otimes dx^j = \sum_{i,j,k,l=0}^{3} F_{ij}\, \frac{\partial x^i}{\partial y^k}\frac{\partial x^j}{\partial y^l}\, dy^k \otimes dy^l$$

and therefore

$$F'_{kl} = \sum_{i,j=0}^{3} F_{ij}\, \frac{\partial x^i}{\partial y^k}\frac{\partial x^j}{\partial y^l} \tag{6.42}$$

In the case of a Lorentz transformation from a system O with coordinates x to a system O' with coordinates y which moving with velocity v along the positive x^1-axis of O we have

$$\Lambda_{ij} = \frac{\partial x^i}{\partial y^k} = \begin{pmatrix} \gamma & \beta & 0 & 0 \\ \beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{6.43}$$

where $\beta = v\gamma/c$. Then Eq.(6.42) becomes

$$F'_{kl} = \sum_{j=0}^{3} (\Lambda^T)_{ki}\, F_{ij}\, \Lambda_{jl}$$

Inserting the matrix (6.43) into this expression the yields

$$
\begin{aligned}
F'_{kl} &= \begin{pmatrix} \gamma & \beta & 0 & 0 \\ \beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & -E_1 & -E_2 & -E_3 \\ E_1 & 0 & B_3 & -B_2 \\ E_2 & -B_3 & 0 & B_1 \\ E_3 & B_2 & -B_1 & 0 \end{pmatrix} \begin{pmatrix} \gamma & \beta & 0 & 0 \\ \beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \\
&= \begin{pmatrix} \gamma & \beta & 0 & 0 \\ \beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} -\beta E_1 & -\gamma E_1 & -E_2 & -E_3 \\ \gamma E_1 & \beta E_1 & B_3 & -B_2 \\ \gamma E_2 - \beta B_3 & \beta E_2 - \gamma B_3 & 0 & B_1 \\ \gamma E_3 + \beta B_2 & \beta E_3 + \gamma B_2 & -B_1 & 0 \end{pmatrix} \\
&= \begin{pmatrix} 0 & -E_1 & -\gamma E_2 + \beta B_3 & -\gamma E_3 - \beta B_2 \\ E_1 & 0 & -\beta E_2 + \gamma B_3 & -\beta E_3 - \gamma B_2 \\ \gamma E_2 - \beta B_3 & \beta E_2 - \gamma B_3 & 0 & B_1 \\ \gamma E_3 + \beta B_2 & \beta E_3 + \gamma B_2 & -B_1 & 0 \end{pmatrix}
\end{aligned}
\tag{6.44}
$$

We therefore see that

$$
\begin{aligned}
E'_1 &= E_1 & B'_1 &= B_1 \\
E'_2 &= \gamma(E_2 - \frac{v}{c}B_3) & B'_2 &= \gamma(B_2 + \frac{v}{c}E_3) \\
E'_3 &= \gamma(E_3 + \frac{v}{c}B_2) & B'_3 &= \gamma(B_3 - \frac{v}{c}E_2)
\end{aligned}
$$

We see that the electric and magnetic field componenst get transformed among each other. We can rewrite these equations as

$$
\begin{aligned}
E'_\parallel &= E'_1 = E_1 \\
B'_\parallel &= B'_1 = B_1 \\
E'_\perp &= \begin{pmatrix} 0 \\ E'_2 \\ E'_3 \end{pmatrix} = \gamma \begin{pmatrix} 0 \\ E_2 \\ E_3 \end{pmatrix} + \frac{\gamma}{c} \begin{pmatrix} v \\ 0 \\ 0 \end{pmatrix} \times \begin{pmatrix} 0 \\ B_2 \\ B_3 \end{pmatrix} \\
B'_\perp &= \begin{pmatrix} 0 \\ B'_2 \\ B'_3 \end{pmatrix} = \gamma \begin{pmatrix} 0 \\ B_2 \\ B_3 \end{pmatrix} - \frac{\gamma}{c} \begin{pmatrix} v \\ 0 \\ 0 \end{pmatrix} \times \begin{pmatrix} 0 \\ E_2 \\ E_3 \end{pmatrix}
\end{aligned}
$$

where we split the electric and magnetic field into components parallel and perpendicular to the the direction of motion of O' with respect to O. Since we can always, for an arbitrary direction of velocity $\mathbf{v}$ of O' with respect to O define the x^1-axis along the direction of motion, we have in general

$$
\mathbf{E}'_\parallel = \mathbf{E}_\parallel \qquad \mathbf{B}'_\parallel = \mathbf{B}_\parallel \tag{6.45}
$$

$$
\mathbf{E}'_\perp = \gamma(\mathbf{E}_\perp + \frac{1}{c}\mathbf{v} \times \mathbf{B}_\perp) \qquad \mathbf{B}'_\perp = \gamma(\mathbf{B}_\perp - \frac{1}{c}\mathbf{v} \times \mathbf{E}_\perp) \tag{6.46}
$$

These are the general transformation laws for the electric and magnetic fields under a Lorentz transformation. There are further two useful invariants that one can construct. These are

$$
\alpha = \sum_{i,j=0}^{3} F_{ij} F^{ij} \qquad \beta = \sum_{i,j=0}^{3} (\star F)_{ij} F^{ij} \tag{6.47}
$$

We have (see Eqs.(3.103) and (3.104)) that

$$F^{ij} = \begin{pmatrix} 0 & E_1 & E_2 & E_3 \\ -E_1 & 0 & B_3 & -B_2 \\ -E_2 & -B_3 & 0 & B_1 \\ -E_3 & B_2 & -B_1 & 0 \end{pmatrix} \qquad (\star F)_{ij} = \begin{pmatrix} 0 & B_1 & B_2 & B_3 \\ -B_1 & 0 & E_3 & -E_2 \\ -B_2 & -E_2 & 0 & E_1 \\ -B_3 & E_2 & -E_1 & 0 \end{pmatrix}$$

which together with Eq.(6.36) gives

$$\alpha = 2\,(\mathbf{B}^2 - \mathbf{E}^2) \quad , \quad \beta = 4\,\mathbf{E}\cdot\mathbf{B} \tag{6.48}$$

This implies that the statement $|\mathbf{E}| > |\mathbf{B}|$ or $|\mathbf{B}| > |\mathbf{E}|$ is a Lorentz invariant. Moreover if $\mathbf{E} = 0$ or $\mathbf{B} = 0$ in one Lorentz frame then $\mathbf{E} \perp \mathbf{B}$ in another Lorentz frame. So if non-zero electric and magnetic fields are not perpendicular then no Lorentz frame can be found in which either the $\mathbf{E}$-field or the $\mathbf{B}$-field vanishes.
All these examples show that electromagnetism is not described by two independent vector fields $\mathbf{E}$ and $\mathbf{B}$. Instead they are described by an anti-symmetric field tensor F with simple transformation properties.

6.5 Particle motion in static fields

6.5.1 The constant electric field

Let us now look at the solution of the Lorentz force Eq.(6.41) for two illustrative cases.
Let a charged particle move in a homogeneous electric field $\mathbf{E} = E_1 e_1 = (E_1, 0, 0)$. The particle has an initial momentum $\mathbf{p} = p_0 e_2$ at $t = 0$ in the x^2-direction.

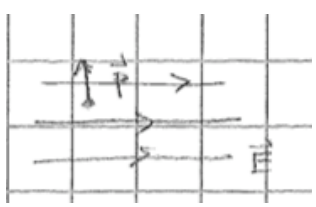

Figure 6.1: Initial momentum p_0 is perpendicular to the direction of the $\mathbf{E}$-field.

According to Eq.(6.41) the equations of motion are

$$\begin{aligned} \frac{\partial}{\partial t}\left(\frac{m}{\sqrt{1-u^2/c^2}}\frac{\partial x^1}{\partial t}\right) &= q\,E_1 \\ \frac{\partial}{\partial t}\left(\frac{m}{\sqrt{1-u^2/c^2}}\frac{\partial x^2}{\partial t}\right) &= 0 \\ \frac{\partial}{\partial t}\left(\frac{m}{\sqrt{1-u^2/c^2}}\frac{\partial x^3}{\partial t}\right) &= 0 \end{aligned}$$

Because $u^2 = (u^1)^2 + (u^2)^2 + (u^3)^2$ appears in every equation the equations are not independent.

Integrating the equations and using the initial condition gives

$$\frac{m}{\sqrt{1-u^2/c^2}}\frac{\partial x^1}{\partial t} = q\,E_1\,t \tag{6.49}$$

$$\frac{m}{\sqrt{1-u^2/c^2}}\frac{\partial x^2}{\partial t} = p_0 \tag{6.50}$$

$$\frac{m}{\sqrt{1-u^2/c^2}}\frac{\partial x^3}{\partial t} = 0$$

The last eqation gives that u^3 is constant, but since $u^3(0) = 0$ we have $u^3(t) = 0$ and hence $u^2 = (u^1)^2 + (u^2)^2$. The Eqs.(6.49) and (6.50) are still coupled. From these two equations we see that

$$\frac{\partial x^1}{\partial t} = \frac{qE_1 t}{p_0}\frac{\partial x^2}{\partial t} \tag{6.51}$$

Squaring Eq.(6.50) gives

$$\left(\frac{\partial x^2}{\partial t}\right)^2 = \left(\frac{p_0}{m}\right)^2 \left(1 - \frac{1}{c^2}\left(\frac{\partial x^1}{\partial t}\right)^2 - \frac{1}{c^2}\left(\frac{\partial x^2}{\partial t}\right)^2\right)$$

and inserting Eq.(6.51) into this equation then gives

$$\left(\frac{\partial x^2}{\partial t}\right)^2 = \left(\frac{p_0}{m}\right)^2 - \left(\frac{p_0}{mc}\right)^2\left(\frac{\partial x^2}{\partial t}\right)^2\left[1 + \left(\frac{qE_1 t}{p_0}\right)^2\right]$$

which yields

$$\left(\frac{\partial x^2}{\partial t}\right)^2 = \frac{(p_0 c^2)^2}{m^2c^4 + p_0^2c^2 + (qcE_1 t)^2}$$

We therefore find

$$u^2(t) = \frac{\partial x^2}{\partial t} = \frac{p_0 c^2}{\sqrt{E_0^2 + (qcE_1 t)^2}} \tag{6.52}$$

where we defined $E_0 = \sqrt{m^2c^4 + p_0^2c^2}$ to be the initial energy of the particle. From Eq.(6.51) we then also find that

$$u^1(t) = \frac{\partial x^1}{\partial t} = \frac{qE_1 c^2 t}{\sqrt{E_0^2 + (qcE_1 t)^2}} \tag{6.53}$$

Both Eq.(6.52) and Eq.(6.53) can now be integrated. If we take $x^1(0) = 0$ and $x^2(0) = 0$ we find

$$x^1(t) = \frac{1}{qE_1}\left[\sqrt{E_0^2 + (qcE_1 t)^2} - E_0\right] \tag{6.54}$$

$$x^2(t) = \frac{p_0 c}{qE_1}\sinh^{-1}\frac{qcE_1 t}{E_0} \tag{6.55}$$

From these results we can see a number of interesting things.First of all, unlike in the Newtonian case, the velocity does not grow linearly with time and does not become arbitrarily large but remains always less than c. Only for times $t \ll E_0/qE_1c$ and $p_0 \ll mc$ we have

$$u^1 \approx \frac{qE_1c^2 t}{E_0} = \frac{qE_1c^2t}{\sqrt{m^2c^4 + p_0^2c^2}} \approx \frac{qE_1}{m}t$$

$$u^2 \approx \frac{p_0c^2}{E_0} \approx \frac{p_0}{m}$$

If we denote $\alpha = qE_1 c/E_0$ then

$$u^1 = \frac{c\alpha t}{\sqrt{1+(\alpha t)^2}} \quad , \quad u^2 = \frac{p_0 c^2}{E_0} \frac{1}{\sqrt{1+(\alpha t)^2}}$$

so as a function of αt we have

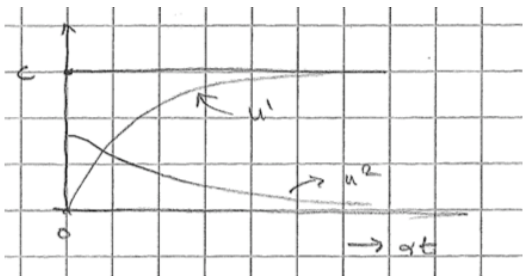

Figure 6.2: The velocities u^1 and u^2 as a function of αt.

We therefore see that, unlike the Newtonian case, that also the velocity u^2 changes with time. Let us further calculate the energy of the particle. Since

$$E = \frac{mc^2}{\sqrt{1-u^2/c^2}}$$

we need to calculate

$$1 - \frac{u^2}{c^2} = 1 - \left[\frac{(\alpha t)^2}{1+(\alpha t)^2} + \left(\frac{p_0 c^2}{E_0} \right)^2 \frac{1}{1+(\alpha t)^2} \right] = \frac{1}{E_0^2} \frac{m^2 c^4}{1+(\alpha t)^2} \tag{6.56}$$

and therefore

$$E = E_0 \sqrt{1+(\alpha t)^2}$$

So for large times the energy grows linearly with time

$$E \approx \alpha E_0 t = qE_1 c\, t \qquad (t \to \infty)$$

We can further calculate the relation between the proper time τ and the time t in the Lorentz frame O. According to Eq.(1.65) and Eq.(6.56) we have

$$\tau(t) = \int_0^t dt' \sqrt{1 - \frac{u^2}{c^2}} = \frac{mc^2}{E_0} \int_0^t dt' \frac{1}{\sqrt{1+(\alpha t')^2}} = \frac{mc}{qE_1} \sinh^{-1}(\alpha t)$$

or equivalently

$$t(\tau) = \frac{E_0}{qE_1 c} \sinh(\frac{qE_1 \tau}{mc}) \tag{6.57}$$

Inserting this expression into Eqs.(6.54) and (6.55) then gives

$$x^1(\tau) = \frac{E_0}{qE_1} \left[\sqrt{1+(\alpha t)^2} - 1 \right] = \frac{E_0}{qE_1} \left[\cosh(\frac{qE_1}{mc} \tau) - 1 \right]$$
$$x^2(\tau) = \frac{p_0}{m} \tau$$

In this way we completely determined the world line of the particle as parametrized by the proper time τ (we take $x^0(\tau) = c\,t(\tau)$ and can take $x^3(\tau) = 0$). We can then directly calculate the four-velocity

$$
\begin{aligned}
v^0(\tau) &= \frac{\partial x^0}{\partial \tau} = c\frac{\partial t}{\partial \tau} = \frac{E_0}{mc}\cosh\left(\frac{qE_1}{mc}\,\tau\right) \\
v^1(\tau) &= \frac{\partial x^1}{\partial \tau} = \frac{E_0}{mc}\sinh\left(\frac{qE_1}{mc}\,\tau\right) \\
v^2(\tau) &= \frac{\partial x^1}{\partial \tau} = \frac{p_0}{m} \\
v^3(\tau) &= 0
\end{aligned}
$$

As a check on our results we can verify that the length of the four-velocity is indeed $-c^2$. We have

$$
\begin{aligned}
g(v,v) = -(v^0)^2 + (v^1)^2 + (v^2)^2 &= \left(\frac{E_0}{mc}\right)^2 \left[-\cosh^2\left(\frac{qE_1}{mc}\,\tau\right) + \sinh^2\left(\frac{qE_1}{mc}\,\tau\right)\right] + \frac{p_0^2}{m^2} \\
&= \frac{p_0^2}{m^2} - \frac{E_0^2}{m^2c^2} = \frac{1}{m^2c^2}(p_0^2c^2 - E_0^2) = -c^2
\end{aligned}
$$

We can further check that

$$
g(v, \frac{Dv}{d\tau}) = -v^0\frac{\partial v^0}{\partial \tau} + v^1\frac{\partial v^1}{\partial \tau} + v^2\frac{\partial v^2}{\partial \tau} = 0
$$

We can finally calculate that

$$
\begin{aligned}
\frac{\partial v^0}{\partial \tau} &= \frac{qE_1}{mc}\frac{E_0}{mc}\sinh\left(\frac{qE_1}{mc}\,\tau\right) = \frac{qE_1}{mc}\,v^1(\tau) \\
\frac{\partial v^1}{\partial \tau} &= \frac{qE_1}{mc}\frac{E_0}{mc}\cosh\left(\frac{qE_1}{mc}\,\tau\right) = \frac{qE_1}{mc}\,v^0(\tau) \\
\frac{\partial v^2}{\partial \tau} &= 0 \\
\frac{\partial v^3}{\partial \tau} &= 0
\end{aligned}
$$

These equations are exactly the equations of motion (6.37) and (6.38) of our problem. We could, of course, also have started by solving these equations and then work out how the equations look in terms of the time-variable t of our Lorentz frame. Let us finally draw the world line on the basis of our equations. We have

$$
\begin{aligned}
x^0(\tau) &= \frac{E_0}{qE_1}\sinh(\frac{qE_1\tau}{mc}) \\
x^1(\tau) &= \frac{E_0}{qE_1}\left[\cosh(\frac{qE_1}{mc}\,\tau) - 1\right] \\
x^2(\tau) &= \frac{p_0}{m}\,\tau
\end{aligned}
$$

and $x^3(\tau) = 0$. This gives the following figure

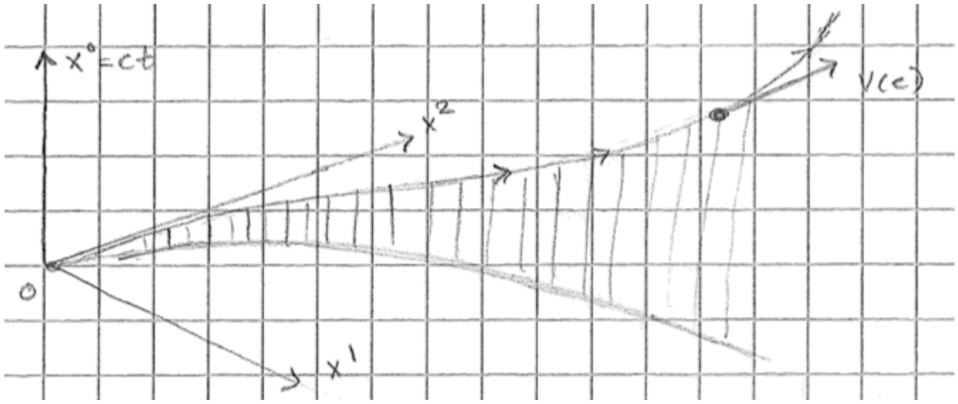

Figure 6.3: The word line of a charged particle accelerated in a static electric field.

6.5.2 The constant magnetic field

The obvious next example is that of a constant magnetic field. Let this field be pointed in the x^3-direction. We therefore take $(B_1, B_2, B_3) = (0, 0, B)$ and $\mathbf{E} = 0$. Let us this time directly solve the equations (6.37) and (6.38) in terms of the proper time. These equations take the form

$$m\frac{\partial v^0}{\partial \tau} = 0 \qquad\qquad m\frac{\partial v^1}{\partial \tau} = \frac{qB}{c}\, v^2(\tau)$$
$$m\frac{\partial v^2}{\partial \tau} = -\frac{qB}{c}\, v^1(\tau) \qquad\qquad m\frac{\partial v^3}{\partial \tau} = 0$$

From these equations we immediately find that v^0 and v^3 must be constant. If we call these constants α and ν then we have

$$v^0 = \frac{\partial x^0}{\partial \tau} = \alpha \quad , \quad v^3 = \frac{\partial x^3}{\partial \tau} = \nu$$

and consequently

$$x^0(\tau) = \alpha\tau \quad , \quad x^3(\tau) = \nu\tau \tag{6.58}$$

where we used the initial conditions $x^0(0) = x^3(0) = 0$. If we define $\omega = qB/mc$ then the equations of motion for v^1 and v^2 can be rewritten as

$$\frac{\partial v^1}{\partial \tau} = \omega\, v^2(\tau)$$
$$\frac{\partial v^2}{\partial \tau} = -\omega\, v^1(\tau)$$

Since

$$\frac{\partial^2 v^1}{\partial \tau^2} = \omega\frac{\partial v^2}{\partial \tau} = -\omega^2\, v^1$$

we find the general solution

$$v^1(\tau) = a\, \cos(\omega\tau) + b\, \sin(\omega\tau)$$

and

$$v^2(\tau) = \frac{1}{\omega}\frac{\partial v^1}{\partial \tau} = -a\, \sin(\omega\tau) + b\, \cos(\omega\tau)$$

These equations can be integrated to give

$$x^1(\tau) = \frac{a}{\omega}\sin(\omega\tau) - \frac{b}{\omega}\cos(\omega\tau) + C$$
$$x^2(\tau) = \frac{a}{\omega}\cos(\omega\tau) - \frac{b}{\omega}\sin(\omega\tau) + D$$

where C and D are integration constants. No generality is lost by letting the particle start in the origin of our coordinate system. Then $x^1(0) = x^2(0) = 0$ gives

$$x^1(0) = C - \frac{b}{\omega} = 0 \quad , \quad x^2(0) = D + \frac{a}{\omega} = 0$$

Further no generality is lost by requiring $v^2(0)$ is zero, as we can always choose the coordinate system in such a way that the initial velocity vector lies in the $x^1 - x^3$ plane. This gives $b = 0$. Collecting our results we then find that

$$x^1(\tau) = \frac{a}{\omega}\sin(\omega\tau)$$
$$x^2(\tau) = \frac{a}{\omega}\left[\cos(\omega\tau) - 1\right]$$

where $a = v^1(0)$. Together with Eq.(6.58) we therefore have

$$x^0(\tau) = \alpha\,\tau$$
$$x^1(\tau) = \frac{v^1(0)}{\omega}\sin(\omega\tau)$$
$$x^2(\tau) = \frac{v^1(0)}{\omega}\left[\cos(\omega\tau) - 1\right]$$
$$x^3(\tau) = v^3(0)\,\tau$$

This determines the world line of the particle in terms of the initial conditions. The constant α can be determined from Eq.(6.6) which gives

$$-c^2 = -\alpha^2 + (v^1(0))^2 + (v^3(0))^2$$

or equivalently

$$\alpha = \sqrt{c^2 + (v^1(0))^2 + (v^3(0))^2} \tag{6.59}$$

Now that we completely determined the world line we can draw the motion of the particle in three dimensions. This is given in the figure

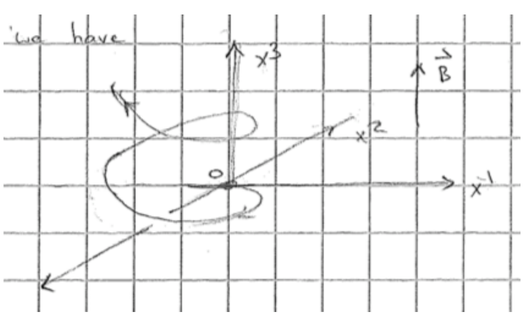

Figure 6.4: Motion of a charged particle in a constant magnetic field.

The particle is spiraling in the positive or negative x^3-direction depending on the sign of $v^3(0)$. Whether the spiraling goes clockwise or counter clockwise depends on the sign of ω or qB. We can change this orientation by either changing the charge of the particle or reverting the direction of the magnetic field.

Chapter 7

Maxwell's equations

We discuss charge conservation and the transformation law of currents and charges under Lorentz transformations. Then we discuss how the Maxwell equations can be written in terms of the covariant derivative of the electromagnetic field tensor. We further show that Maxwell's equations can also be derived from exterior calculus and show the connection between the two formulations. Finally we discuss the form of Maxwell's equations in general orthogonal coordinate systems.

7.1 Currents and conservation of charge

In the previous Chapter we saw how electric and magnetic fields are transformed under Lorentz transformations.Let us now see how charge and current densities are transformed. We start by deriving the charge conservation law in differential form. The basic assumption is the well-verified experimental fact that in any process charge is conserved. Another fact that we will use is that the total charge of an object is independent of its state of motion, i.e. the same for any observer.
Consider, in a given Lorentz frame, a continuous charge distribution $\rho(\mathbf{x}, t)$ moving with a three-dimensional velocity field $\mathbf{v}(\mathbf{x}, t)$. Consider further a given volume V at rest with respect to our reference frame and we measure the charge that flows into this volume.

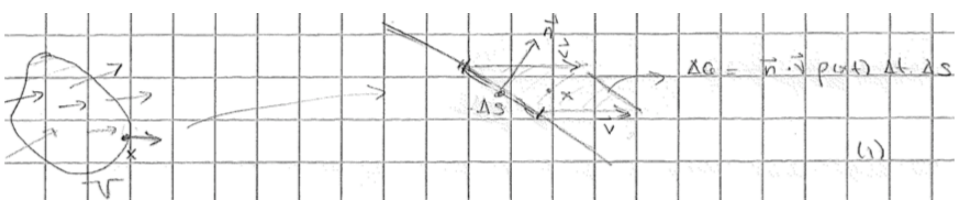

Figure 7.1: Flow of charge through a volume

The amount of charge ΔQ flowing in a time Δt through a surface element ΔS is given by

$$\Delta Q = \mathbf{n} \cdot \mathbf{v}(\mathbf{x}, t)\rho(\mathbf{x}, t)\Delta t \, \Delta S$$

Therefore the total charge per unit time flowing through the surface S of V is given by the surface integral

$$\frac{\partial Q}{\partial t} = \int_S dS \, \mathbf{n} \cdot \mathbf{v}(\mathbf{x}, t)\rho(\mathbf{x}, t)$$

If charge is conserved then the charge flowing out is equal to the change in volume charge which is given by

$$\frac{\partial Q}{\partial t} = -\frac{\partial}{\partial t}\int_V d\mathbf{x}\,\rho(\mathbf{x},t)$$

From these two equations we find that

$$0 = \int_V d\mathbf{x}\,\frac{\partial}{\partial t}\rho(\mathbf{x},t) + \int_S dS\,\mathbf{n}\cdot\mathbf{v}(\mathbf{x},t)\rho(\mathbf{x},t) = \int_V d\mathbf{x}\left(\frac{\partial}{\partial t}\rho(\mathbf{x},t) + \nabla\cdot(\rho(\mathbf{x},t)\mathbf{v}(\mathbf{x},t))\right)$$

where in the last step we used Gauss' law to convert a surface integral to a volume integral. Since this relation is true for any volume V we must have

$$0 = \frac{\partial}{\partial t}\rho(\mathbf{x},t) + \nabla\cdot(\rho(\mathbf{x},t)\mathbf{v}(\mathbf{x},t)) \tag{7.1}$$

This is the equation for the conservation of charge in differential form. The quantity

$$\mathbf{j}(\mathbf{x},t) = \rho(\mathbf{x},t)\mathbf{v}(\mathbf{x},t) \tag{7.2}$$

is called the *current density*. Let us now see charge and current densities look in different reference frames. Let us, for simplicity, take a homogeneous charge distribution ρ_0 in a wire of length L_0 at rest in reference frame O.

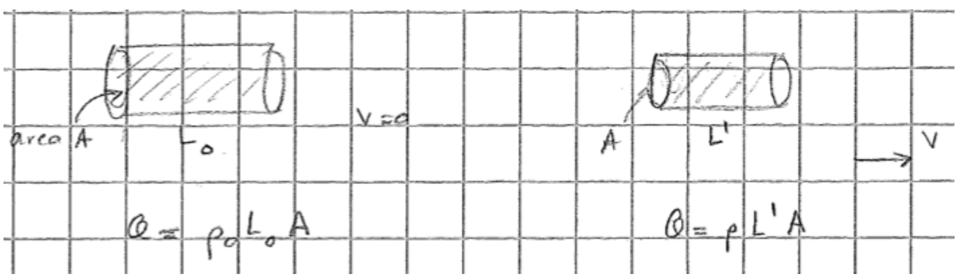

Figure 7.2: Charge in a wire

In a system O' moving at parallel velocity v w.r.t. the wire the length of the wire is found Lorentz contracted to

$$L' = L_0\sqrt{1-\frac{v^2}{c^2}}$$

Since we assumed that the total charge of an object is independent of its state of motion (an experimental fact) then we must have that the charge density ρ in the moving frame satisfies

$$\rho_0 L_0 = \rho L' = \rho L_0\sqrt{1-\frac{v^2}{c^2}}$$

and therefore that

$$\rho = \frac{\rho_0}{\sqrt{1-\frac{v^2}{c^2}}} = \gamma\,\rho_0$$

So if ρ_0 is the charge density of an object at rest then Eq.(7.3) gives the charge density of the object at velocity v. In that frame the current density is then given by

$$\mathbf{j} = \rho\,\mathbf{v} = \frac{\rho_0\mathbf{v}}{\sqrt{1-\frac{v^2}{c^2}}} = \gamma\,\rho_0\mathbf{v}$$

From this analysis we see that it is natural define a current four-vector j in terms of the four-velocity $v = \gamma(c, \mathbf{v})$ by

$$j = \rho_0\, v = \rho_0 \gamma(c, \mathbf{v}) = (c\,\rho, \mathbf{j}) \tag{7.3}$$

In the example we used uniform charge densities and currents but it is clear that this can be readily generalized to non-uniform ones. We define the four-vector

$$j(x) = (c\rho(\mathbf{x}, t), \mathbf{j}(\mathbf{x}, t)) = (j^0(x), j^1(x), j^2(x), j^3(x)) \tag{7.4}$$

where $x = (x^0, x^1, x^2, x^3)$ and $x^0 = ct$. This vector satisfies the condition

$$\sum_{k=0}^{3} \frac{\partial j^k}{\partial x^k} = \frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{j} = 0 \tag{7.5}$$

which is just the condition for conservation of charge of Eq.(7.1). Under a coordinate to a new coordinate system y transformation j transforms as a vector, i.e.

$$j'^k(y) = \sum_{l=0}^{3} \frac{\partial y^k}{\partial x^l}\, j^l(x) \tag{7.6}$$

In case y^k are coordinates in a Lorentz frame O' moving with respect to our original frame O with constant speed v along the x^1 axis the transformation matrix is given by Eq.(6.23). Using explicitly the components of j of Eq.(7.4) then gives the transformation law of charge densities and three-dimensional currents. We find

$$j'^1(y) = \frac{j^1(x) - v\,\rho(x)}{\sqrt{1 - \frac{v^2}{c^2}}} \tag{7.7}$$

$$\rho'(y) = \frac{\rho(x) - v\, j^1(x)/c^2}{\sqrt{1 - \frac{v^2}{c^2}}} \tag{7.8}$$

$$j'^2(y) = j^2(x) \quad , \quad j'^3(y) = j^3(x) \tag{7.9}$$

where the left hand sides are regarded as functions of y by regarding $x(y)$ as a function of y in the arguments on the right hand sides. We can check that this equation is consistent with the formula for addition of velocities of Eqs.(1.47)-(1.49). Let $\mathbf{j}'(y)/\rho'(y) = \mathbf{u}'(y)$ be the velocity field of the charged fluid in system O' and $\mathbf{j}(x)/\rho(x) = \mathbf{u}(x)$ be the velocity field of the charged fluid in system O. Then from Eqs.(7.7) and (7.8) it follows that

$$u'^1(y) = \frac{j'^1(y)}{\rho'(y)} = \frac{j^1(x) - v\,\rho(x)}{\rho(x) - v\, j^1(x)/c^2} = \frac{j^1(x)/\rho(x) - v}{1 - v\,(j^1(x)/\rho(x))/c^2} = \frac{u^1(x) - v}{1 - v\, u^1(x)/c^2}$$

which is consistent with Eq.(1.47). Similarly for the component u'^2 we find

$$\begin{aligned} u'^2(y) = \frac{j'^2(y)}{\rho'(y)} &= \frac{j^2(x)}{\rho(x) - v\, j^1(x)/c^2} \sqrt{1 - \frac{v^2}{c^2}} \\ &= \frac{j^2(x)/\rho(x)}{1 - v\,(j^1(x)/\rho(x))/c^2} \sqrt{1 - \frac{v^2}{c^2}} = \frac{u^2(x)\sqrt{1 - \frac{v^2}{c^2}}}{1 - v\, u^2(x)/c^2} \end{aligned}$$

which is consistent with Eq.(1.48). The equation for $u'^3(y)$ is clearly identical.

7.2 The Maxwell equations

Charges and currents are the sources of the electromagnetic fields. We therefore want to make a connection between the four-current j and the electromagnetic field tensor F. It is known experimentally that this relation is linear. The fields produced are proportional to the strengths of the currents and field produced by a superposition of currents is equal to the superposition of the fields produced by each current separately. If we denote mapping $j \mapsto F$ by $F[j]$, then we have

$$F[\alpha\, j_1 + \beta\, j_2] = \alpha\, F[j_1] + \beta\, F[j_2] \tag{7.10}$$

where α, β are some real numbers. We therefore want a linear relation between the rank two tensor $F_{\mu\nu}$ and the four-current j^μ. One of the simplest ways to produce a vector out of a second order covariant tensor is to raise the indices and to perform a contraction on the covariant derivative. So we try

$$\sum_{\nu=0}^{3} F^{\mu\nu}_{;\nu} = \alpha\, j^\mu \tag{7.11}$$

where α is a constant to be determined. Let us see what we get. The contravariant form of F was in Eq.(3.103) to be

$$F^{\mu\nu} = \begin{pmatrix} 0 & E_1 & E_2 & E_3 \\ -E_1 & 0 & B_3 & -B_2 \\ -E_2 & -B_3 & 0 & B_1 \\ -E_3 & B_2 & -B_1 & 0 \end{pmatrix} \tag{7.12}$$

Let us further take the case that we use standard Minkowskian coordinates, such that

$$F^{\mu\nu}_{;\nu} = \frac{\partial F^{\mu\nu}}{\partial x^\nu}$$

then our guess Eq.(7.11) implies that

$$\begin{aligned}
\frac{\partial E_1}{\partial x^1} + \frac{\partial E_2}{\partial x^2} + \frac{\partial E_3}{\partial x^3} &= \alpha j^0 \\
-\frac{\partial E_1}{\partial x^0} + \frac{\partial B_3}{\partial x^2} - \frac{\partial B_2}{\partial x^3} &= \alpha j^1 \\
-\frac{\partial E_2}{\partial x^0} - \frac{\partial B_3}{\partial x^1} + \frac{\partial B_1}{\partial x^3} &= \alpha j^2 \\
-\frac{\partial E_3}{\partial x^0} + \frac{\partial B_2}{\partial x^1} - \frac{\partial B_1}{\partial x^2} &= \alpha j^3
\end{aligned}$$

We see that these equations can be rewritten as

$$\frac{\partial E_1}{\partial x^1} + \frac{\partial E_2}{\partial x^2} + \frac{\partial E_3}{\partial x^3} = \alpha j^0 = \alpha c \rho$$

$$-\frac{1}{c}\frac{\partial}{\partial t}\begin{pmatrix} E_1 \\ E_2 \\ E_3 \end{pmatrix} + \begin{pmatrix} \frac{\partial}{\partial x^1} \\ \frac{\partial}{\partial x^2} \\ \frac{\partial}{\partial x^3} \end{pmatrix} \times \begin{pmatrix} B_1 \\ B_2 \\ B_3 \end{pmatrix} = \alpha \begin{pmatrix} j^1 \\ j^2 \\ j^3 \end{pmatrix}$$

or equivalently as

$$\begin{aligned}
\nabla \cdot \mathbf{E} &= \alpha c\, \rho \\
-\frac{1}{c}\frac{\partial \mathbf{E}}{\partial t} + \nabla \times \mathbf{B} &= \alpha\, \mathbf{j}
\end{aligned}$$

We recognize these equations as two of Maxwell's equations when we put $\alpha = 4\pi/c$ (in Gaussian units). Then we obtain

$$\nabla \cdot \mathbf{E} = 4\pi\,\rho \tag{7.13}$$

$$-\frac{1}{c}\frac{\partial \mathbf{E}}{\partial t} + \nabla \times \mathbf{B} = \frac{4\pi}{c}\,\mathbf{j} \tag{7.14}$$

We therefore discover two of Maxwell's equations from the relation

$$\sum_{\nu=0}^{3} F^{\mu\nu}_{;\nu} = \frac{4\pi}{c}\,j^{\mu} \tag{7.15}$$

Equations (7.13) and (7.14) do not determine $\mathbf{E}$ and $\mathbf{B}$ uniquely. It is, for instance, readily seen that they determine $\mathbf{B}$ up to a gradient of a scalar function since the curl of gradient is zero. This non-uniqueness can be resolved if we have another equations where $\mathbf{E}$ appears as a curl and $\mathbf{B}$ as a divergence. A simple guess is simply to interchange $\mathbf{E}$ and $\mathbf{B}$ in Eqs.(7.13) and (7.14) which would give the required equations. However, since (as far as experimental evidence goes) there are no magnetic charges and currents we should put $j = 0$ in such an equation. The question therefore is if we can manipulate F in such a way that the roles of the electric and magnetic fields are interchanged. But, as we learned from Eq.(3.104), it is exactly the Hodge dual of F that achieves this. If we raise the indices on $\star F$ we have

$$(\star F)^{\mu\nu} = \begin{pmatrix} 0 & -B_1 & -B_2 & -B_3 \\ B_1 & 0 & E_3 & -E_2 \\ B_2 & -E_3 & 0 & E_1 \\ B_3 & E_2 & -E_1 & 0 \end{pmatrix} \tag{7.16}$$

and we then require

$$\sum_{\nu=0}^{3} (\star F)^{\mu\nu}_{;\nu} = 0 \tag{7.17}$$

This then yields the equations

$$\begin{aligned} -\frac{\partial B_1}{\partial x^1} - \frac{\partial B_2}{\partial x^2} - \frac{\partial B_3}{\partial x^3} &= 0 \\ \frac{\partial B_1}{\partial x^0} + \frac{\partial E_3}{\partial x^2} - \frac{\partial E_2}{\partial x^3} &= 0 \\ \frac{\partial B_2}{\partial x^0} - \frac{\partial E_3}{\partial x^1} + \frac{\partial E_1}{\partial x^3} &= 0 \\ \frac{\partial B_3}{\partial x^0} + \frac{\partial E_2}{\partial x^1} - \frac{\partial E_1}{\partial x^2} &= 0 \end{aligned}$$

or equivalently

$$\nabla \cdot \mathbf{B} = 0 \tag{7.18}$$

$$\frac{1}{c}\frac{\partial \mathbf{B}}{\partial t} + \nabla \times \mathbf{E} = 0 \tag{7.19}$$

These are the remaining two Maxwell equations needed to obtain a unique solution for the $\mathbf{E}$ and $\mathbf{B}$ fields from the current. Together with the result of Chapter 6 for the Lorentz force we thus arrived at the Maxwell and Lorentz laws which we summarize together as

$$\nabla \cdot \mathbf{E} = 4\pi\,\rho \tag{7.20}$$
$$-\frac{1}{c}\frac{\partial \mathbf{E}}{\partial t} + \nabla \times \mathbf{B} = \frac{4\pi}{c}\,\mathbf{j} \tag{7.21}$$
$$\nabla \cdot \mathbf{B} = 0 \tag{7.22}$$
$$\frac{1}{c}\frac{\partial \mathbf{B}}{\partial t} + \nabla \times \mathbf{E} = 0 \tag{7.23}$$
$$\frac{\partial}{\partial t}\Big(\frac{m\mathbf{v}}{\sqrt{1-\frac{v^2}{c^2}}}\Big) = q(\mathbf{E} + \frac{1}{c}\,\mathbf{v} \times \mathbf{B}) \tag{7.24}$$

These equations summarize everything there is to know about classical electromagnetism and what was known about classical electromagnetism after Einstein published his paper on special relativity. In terms of the electromagnetic field tensor these equations can be summarized in a much more symmetric way as

$$\sum_{\nu=0}^{3} F^{\mu\nu}_{;\nu} = \frac{4\pi}{c}\,j^{\mu} \tag{7.25}$$
$$\sum_{\nu=0}^{3} (\star F)^{\mu\nu}_{;\nu} = 0 \tag{7.26}$$
$$m\left(\frac{Dv}{dt}\right)_{\mu} = \frac{q}{c}\sum_{\nu=0}^{3} F_{\mu\nu}\,v^{\nu} \tag{7.27}$$

What remains is to get a deeper understanding of these equations and to work them out in various circumstances. As these stand, Eqs.(7.25)-(7.27) are already quite beautiful as a physical theory. However, with a little extra work we can get even more compact equations, and some additional geometric insight. Towards the end of these Lectures we will finally see that F appears as the curvature of a so-called gauge connection.

7.3 Charge conservation revisited

We have seen that Maxwell's equations can be written as a covariant divergence. The same is true for the charge conservation law of Eq.(7.5). We wrote it in terms of standard Minkowskian coordinates, but it is clear that in general coordinates we have

$$0 = \sum_{\mu=0}^{3} j^{\mu}_{;\mu} \tag{7.28}$$

This, in fact, follows directly from Eq.(7.25), since

$$0 = \sum_{\nu,\mu}^{3} (F^{\mu\nu}_{;\nu})_{;\mu} \tag{7.29}$$

We may expect this to follow from the anti-symmetry of F together with the commutativity of differentiation. However, in general covariant derivatives do not commute except when we are dealing with a flat space as will be discussed in the next Chapter, so we have to be a bit careful. However, we can avoid using with Christoffel symbols when are dealing with divergences. Let us analyze this a bit further. The covariant divergence of a vector j is given by (in principle we could now consider again the case of arbitrary number of dimensions n)

$$\sum_\mu j^\mu_{;\mu} = \sum_\mu \frac{\partial j^\mu}{\partial x^\mu} + \sum_{\mu,\rho} j^\rho \Gamma^\mu_{\mu\rho} \tag{7.30}$$

The last term in this equation involves a Christoffel symbol with two indices contracted. From the definitions (4.12) and (4.15) of the Christoffel symbols we have

$$\sum_k \Gamma^k_{kl} = \sum_{k,m} \frac{g^{km}}{2} \left(\frac{\partial g_{mk}}{\partial x^l} + \frac{\partial g_{ml}}{\partial x^k} - \frac{\partial g_{kl}}{\partial x^m} \right) = \sum_{k,m} \frac{g^{km}}{2} \frac{\partial g_{mk}}{\partial x^l} \tag{7.31}$$

We can derive a useful expression for the last term in this equation using the determinant of the metric. Let us take the n-dimensional case again. The determinant of the metric is given by

$$g = \det(g_{ik}) = \sum_{i_1 \dots i_n}^n \epsilon_{i_1 \dots i_n} \, g_{1i_1} \dots g_{ni_n}$$

and we have

$$\frac{\partial g}{\partial x^l} = \sum_{k=1}^n \sum_{i_1 \dots i_n}^n \epsilon_{i_1 \dots i_n} \, g_{1i_1} \dots \frac{\partial g_{ki_k}}{\partial x^l} \dots g_{ni_n} \tag{7.32}$$

If we insert

$$\frac{\partial g_{ki_k}}{\partial x^l} = \sum_m^n \frac{\partial g_{km}}{\partial x^l} \delta^m_{i_k} = \sum_{m,r}^n \frac{\partial g_{km}}{\partial x^l} g^{mr} g_{ri_k}$$

into Eq.(7.32), then if $r \neq k$ the product in Eq.(7.32) contacts the term $g_{ri_k} g_{ri_r}$ which is symmetric in i_k and i_r. Due to the anti-symmetry of ϵ such a term does not contribute to the total sum. Therefore only the term with $r = k$ survives, which gives

$$\frac{\partial g}{\partial x^l} = \sum_{m,k}^n \frac{\partial g_{km}}{\partial x^l} g^{mk} \, g$$

and hence

$$\sum_{m,k}^n \frac{\partial g_{km}}{\partial x^l} g^{mk} = \frac{1}{g} \frac{\partial g}{\partial x^l}$$

From Eq.(7.31) we then see that

$$\sum_k^n \Gamma^k_{kl} = \frac{1}{2g} \frac{\partial g}{\partial x^l} = \frac{1}{\sqrt{|g|}} \frac{\partial}{\partial x^l} \sqrt{|g|}$$

Consequently the covariant divergence in Eq.(7.30) becomes

$$\sum_\mu j^\mu_{;\mu} = \sum_\mu \frac{\partial j^\mu}{\partial x^\mu} + \sum_\rho j^\rho \frac{1}{\sqrt{|g|}} \frac{\partial}{\partial x^\rho} \sqrt{|g|} = \frac{1}{\sqrt{|g|}} \sum_\mu \frac{\partial}{\partial x^\mu} \left(\sqrt{|g|} \, j^\mu \right)$$

It is clear that for the case of a Minkowskian metric $|g| = 1$ and we recover Eq.(7.5). Let us now see what we get for the contravariant electromagnetic field tensor $F^{\mu\nu}$. We have

$$F^{\mu\nu}_{;\rho} = \frac{\partial F^{\mu\nu}}{\partial x^\rho} + \sum_\sigma F^{\sigma\nu} \Gamma^\mu_{\rho\sigma} + \sum_\sigma F^{\mu\sigma} \Gamma^\nu_{\rho\sigma}$$

and therefore

$$\begin{aligned}\sum_\nu F^{\mu\nu}_{;\nu} &= \sum_\nu \frac{\partial F^{\mu\nu}}{\partial x^\nu} + \sum_{\sigma,\nu} F^{\sigma\nu}\Gamma^\mu_{\nu\sigma} + \sum_{\sigma,\nu} F^{\mu\sigma}\Gamma^\nu_{\nu\sigma} \\ &= \sum_\nu \frac{\partial F^{\mu\nu}}{\partial x^\nu} + \sum_\sigma F^{\mu\sigma}\frac{1}{\sqrt{|g|}}\frac{\partial}{\partial x^\sigma}\sqrt{|g|} = \frac{1}{\sqrt{|g|}}\sum_\nu \frac{\partial}{\partial x^\nu}\left(\sqrt{|g|}\,F^{\mu\nu}\right)\end{aligned}$$

where we used that the second term of the equal sign in the first line vanishes due to the anti-symmetry of F and the symmetry of the Christoffel symbol. From this expression we see that Eq.(7.25) can be rewritten as

$$\frac{4\pi}{c}\,j^\mu = \frac{1}{\sqrt{|g|}}\sum_{\nu=0}^{3}\frac{\partial}{\partial x^\nu}\left(\sqrt{|g|}\,F^{\mu\nu}\right) \tag{7.33}$$

But this implies that

$$\frac{4\pi}{c}\sum_{\mu=0}^{3} j^\mu_{;\mu} = \frac{4\pi}{c}\frac{1}{\sqrt{|g|}}\sum_\mu \frac{\partial}{\partial x^\mu}\left(\sqrt{|g|}\,j^\mu\right) = \frac{1}{\sqrt{|g|}}\sum_{\mu,\nu=0}^{3}\frac{\partial}{\partial x^\mu \partial x^\nu}\left(\sqrt{|g|}\,F^{\mu\nu}\right) = 0$$

as a consequence of the anti-symmetry of F. Therefore Eq.(7.25) automatically implies that the covariant divergence of the current vanishes and therefore that charge is conserved.
We saw that the expression for the covariant divergence is simplified due to the anti-symmetry of F. This happens for anti-symmetric tensors in general. Let $F^{i_1\ldots i_p}$ be such a tensor then from Eq.(5.49) we find that

$$\begin{aligned}\sum_k F^{i_1\ldots i_{p-1}k}_{;k} =& \sum_k \frac{\partial F^{i_1\ldots i_{p-1}k}}{\partial x^k} + \sum_{k,l} F^{li_2\ldots i_{p-1}k}\Gamma^{i_1}_{kl} + \ldots + \sum_{k,l} F^{i_1\ldots i_{p-2}lk}\Gamma^{i_{p-1}}_{kl} \\ &+ \sum_{k,l} F^{i_1\ldots i_{p-1}l}\Gamma^k_{kl} \\ =& \sum_k \frac{\partial F^{i_1\ldots i_{p-1}k}}{\partial x^k} + \sum_l F^{i_1\ldots i_{p-1}l}\frac{1}{\sqrt{|g|}}\frac{\partial}{\partial x^l}\sqrt{|g|} \\ =& \frac{1}{\sqrt{|g|}}\sum_k \frac{\partial}{\partial x^k}\left(\sqrt{|g|}\,F^{i_1\ldots i_{p-1}k}\right)\end{aligned} \tag{7.34}$$

where all, except the first and the last, terms after the first equal sign vanish due to the anti-symmetry of F and the symmetry of the Christoffel symbols.

7.4 The exterior derivative

We have seen that in electromagnetism the electromagnetic field tensor F plays a central role. A key property of this tensor is its anti-symmetry $F_{\mu\nu} = -F_{\nu\mu}$. It turns out that anti-symmetric tensors have special properties that allow for the development of an elegant calculus which not only gives insight into the origin of the standard vector operations such as div, curl and the Laplacian, but also allows us to write Maxwell's equations in an even more compact way. The notation will be very useful later when we discuss the derivation of Maxwell's equations from an action principle.
Let us start with an example. Consider a second order covariant tensor F

$$F = \sum_{i,j}^{n} F_{ij}\,dx^i \otimes dx^j$$

In Chapter 5 we saw that we can assign a third order tensor ∇F to F by the operation

$$\nabla F = \sum_{i,j,k}^{n} F_{ij;k}\, dx^k \otimes dx^i \otimes dx^j$$

where

$$F_{ij;k} = \frac{\partial F_{ij}}{\partial x^k} - \sum_{l}^{n} F_{lj} \Gamma^l_{ki} - \sum_{l}^{n} F_{il} \Gamma^l_{kj} \tag{7.35}$$

The last two terms with the Christoffel symbols are very important for making ∇F transform as a tensor. However, if F is anti-symmetric a description can be given for a new type of derivative called dF that maps F to a new anti-symmetric tensor dF which does not require the introduction of Christoffel symbols or a metric. Let us see what happens if we forget about the Christoffel symbols and simply define

$$\tilde{\nabla} F = \sum_{i,j,k}^{n} \frac{\partial F_{ij}}{\partial x^k}\, dx^k \otimes dx^i \otimes dx^j$$

If we go to a new coordinate system y then this equation becomes

$$\tilde{\nabla} F = \sum_{i,j,k,p,q,r}^{n} \frac{\partial F_{ij}}{\partial x^k} \frac{\partial x^k}{\partial y^p} \frac{\partial x^i}{\partial y^q} \frac{\partial x^i}{\partial y^r}\, dy^p \otimes dy^q \otimes dy^r \tag{7.36}$$

We could also first have carried out the coordinate transformation and then apply $\tilde{\nabla}$. We then have

$$F = \sum_{i,j,q,r}^{n} F_{ij} \frac{\partial x^i}{\partial y^q} \frac{\partial x^j}{\partial y^r}\, dy^q \otimes dy^r$$

and

$$\begin{aligned} \tilde{\nabla} F &= \sum_{i,j,p,q,r}^{n} \frac{\partial}{\partial y^p} \left(F_{ij} \frac{\partial x^i}{\partial y^q} \frac{\partial x^j}{\partial y^r} \right) dy^p \otimes dy^q \otimes dy^r \\ &= \sum_{i,j,k,p,q,r} \frac{\partial F_{ij}}{\partial x^k} \frac{\partial x^k}{\partial y^p} \frac{\partial x^i}{\partial y^q} \frac{\partial x^i}{\partial y^r}\, dy^p \otimes dy^q \otimes dy^r \\ &\quad + \sum_{i,j,p,q,r}^{n} F_{ij} \left(\frac{\partial^2 x^i}{\partial y^p \partial y^q} \frac{\partial x^j}{\partial y^r} + \frac{\partial x^i}{\partial y^q} \frac{\partial^2 x^j}{\partial y^p \partial y^r} \right) dy^p \otimes dy^q \otimes dy^r \end{aligned}$$

When we compare this expression to Eq.(7.36) we see that $\tilde{\nabla} F$ has no invariant meaning since the last two terms prevent $\tilde{\nabla} F$ from transforming as a tensor. This is exactly why we needed the Christoffel symbols in the first place. We note, however, that the two unwanted terms are symmetric in p and q and in p and r respectively. The undesirable terms therefore disappear when we make the replacement

$$dy^p \otimes dy^q \otimes dy^r \quad \rightarrow \quad dy^p \wedge dy^q \wedge dy^r$$

as the latter term is anti-symmetric. Therefore, for an anti-symmetric tensor F given by

$$F = \sum_{i,j}^{n} F_{ij}\, dx^i \wedge dx^j$$

we can define

$$dF = \sum_{i,j,k}^{n} \frac{\partial F_{ij}}{\partial x^k} \, dx^k \wedge dx^i \wedge dx^j$$

and find that it this expression does have an invariant meaning. It is clear that this works for general anti-symmetric covariant tensors (also known as p-forms). For a general anti-symmetric F

$$F = \sum_{i_1 \ldots i_p}^{n} F_{i_1 \ldots i_p} \, dx^{i_1} \wedge \ldots \wedge dx^{i_p}$$

we define

$$dF = \sum_{i_1 \ldots i_p, k}^{n} \frac{\partial F_{i_1 \ldots i_p}}{\partial x^k} \, dx^k \wedge dx^{i_1} \wedge \ldots \wedge dx^{i_p} \tag{7.37}$$

We therefore see that the operation d, which is called the exterior derivative, is a mapping

$$d : \Omega^p \mapsto \Omega^{p+1}$$

from p-forms to $(p+1)$-forms. If we define an ordinary function as a 0-form then we see that Eq.(7.37) generalizes Eq.(3.122), i.e.

$$df = \sum_{j}^{n} \frac{\partial f}{\partial x^j} \, dx^j$$

which is mapping from 0-forms to 1-forms. Let us derive some useful properties of the exterior derivative. First of all, it is clear that the operation is linear

$$d(\omega + \mu) = d\omega + d\mu \qquad \omega, \mu \in \Omega^p \tag{7.38}$$

$$d(c\omega) = c \, d\omega \qquad c \in \mathbb{R} \tag{7.39}$$

Then if $\omega \in \Omega^p$ and $\mu \in \Omega^q$ then

$$\begin{aligned}
d(\omega \wedge \mu) =& d\left(\sum_{i_1 \ldots i_p, j_1 \ldots j_q}^{n} \omega_{i_1 \ldots i_p} \mu_{j_1 \ldots j_q} \, dx^{i_1} \wedge \ldots \wedge dx^{i_p} \wedge dx^{j_1} \wedge \ldots \wedge dx^{j_q} \right) \\
=& \sum_{i_1 \ldots i_p, j_1 \ldots j_q, k}^{n} \left(\frac{\partial \omega_{i_1 \ldots i_p}}{\partial x^k} \mu_{j_1 \ldots j_q} + \omega_{i_1 \ldots i_p} \frac{\partial \mu_{j_1 \ldots j_q}}{\partial x^k} \right) dx^k \wedge dx^{i_1} \wedge \ldots dx^{j_q} \\
=& \sum_{i_1 \ldots i_p, j_1 \ldots j_q, k}^{n} \frac{\partial \omega_{i_1 \ldots i_p}}{\partial x^k} \mu_{j_1 \ldots j_q} \, dx^k \wedge dx^{i_1} \wedge \ldots \wedge dx^{j_q} \\
&+ \sum_{i_1 \ldots i_p, j_1 \ldots j_q, k}^{n} (-1)^p \, \omega_{i_1 \ldots i_p} \frac{\partial \mu_{j_1 \ldots j_q}}{\partial x^k} \, dx^{i_1} \wedge \ldots dx^{i_p} \wedge dx^k \wedge dx^{j_1} \wedge \ldots \wedge dx^{j_q} \\
=& \; d\omega \wedge \mu + (-1)^p \, \omega \wedge d\mu
\end{aligned}$$

where a factor $(-1)^p$ appeared in the last term as we moved dx^k over p positions. So we find

$$d(\omega \wedge \mu) = d\omega \wedge \mu + (-1)^p \, \omega \wedge d\mu \tag{7.40}$$

Finally we have that

$$d(d\omega) = d\left(\sum_{i_1\dots i_p,k}^{n} \frac{\partial \omega_{i_1\dots i_p}}{\partial x^k}\, dx^k \wedge dx^{i_1} \wedge \dots \wedge dx^{i_p}\right)$$
$$= \sum_{i_1\dots i_p,k,l}^{n} \frac{\partial^2 \omega_{i_1\dots i_p}}{\partial x^k \partial x^l}\, dx^l \wedge dx^k \wedge dx^{i_1} \wedge \dots \wedge dx^{i_p} = 0$$

as a consequence of the symmetry in k and l of the second derivative. We thus find that

$$d^2\omega = 0 \tag{7.41}$$

for all $\omega \in \Omega^p$. The exterior derivative incorporates all well-known operations of vector analysis as special cases. Let us give a few examples.

Gradient. If $f \in \Omega^0$ we have

$$df = \sum_{j}^{n} \frac{\partial f}{\partial x^j}\, dx^j = \left(\frac{\partial f}{\partial x^1}, \dots, \frac{\partial f}{\partial x^n}\right) \tag{7.42}$$

which gives the standard gradient vector.

Curl. If $\omega \in \Omega^1$ then

$$\omega = \sum_{j}^{n} \omega_j\, dx^j = (\omega_1, \dots, \omega_n)$$

and

$$d\omega = \sum_{j,k}^{n} \frac{\partial \omega_j}{\partial x^k}\, dx^k \wedge dx^j = \sum_{k<j}^{n} \left(\frac{\partial \omega_j}{\partial x^k} - \frac{\partial \omega_k}{\partial x^j}\right) dx^k \wedge dx^j$$

In three dimensions this gives

$$d\omega = \left(\frac{\partial \omega_2}{\partial x^1} - \frac{\partial \omega_1}{\partial x^2}\right) dx^1 \wedge dx^2 + \left(\frac{\partial \omega_3}{\partial x^1} - \frac{\partial \omega_1}{\partial x^3}\right) dx^1 \wedge dx^3 + \left(\frac{\partial \omega_3}{\partial x^2} - \frac{\partial \omega_2}{\partial x^3}\right) dx^2 \wedge dx^3$$

If x^i are Cartesian coordinates and we have a standard Euclidean metric of the form

$$g = dx^1 \otimes dx^1 + dx^2 \otimes dx^2 + dx^3 \otimes dx^3$$

then it follows from Eq.(3.94) that

$$\star\, dx^1 \wedge dx^2 = dx^3 \qquad \star\, dx^1 \wedge dx^3 = -dx^2 \qquad \star\, dx^2 \wedge dx^3 = dx^1$$

and therefore

$$\star\, d\omega = \left(\frac{\partial \omega_3}{\partial x^2} - \frac{\partial \omega_2}{\partial x^3}\right) dx^1 + \left(\frac{\partial \omega_1}{\partial x^3} - \frac{\partial \omega_3}{\partial x^1}\right) dx^2 + \left(\frac{\partial \omega_2}{\partial x^1} - \frac{\partial \omega_1}{\partial x^2}\right) dx^3$$
$$= \left(\frac{\partial \omega_3}{\partial x^2} - \frac{\partial \omega_2}{\partial x^3}, \frac{\partial \omega_1}{\partial x^3} - \frac{\partial \omega_3}{\partial x^1}, \frac{\partial \omega_2}{\partial x^1} - \frac{\partial \omega_1}{\partial x^2}\right) = \begin{pmatrix} \frac{\partial}{\partial x^1} \\ \frac{\partial}{\partial x^2} \\ \frac{\partial}{\partial x^3} \end{pmatrix} \times \begin{pmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \end{pmatrix}$$

which is the familiar curl of a vector field.

Divergence. Let ω be a 1-form in a three-dimensional Euclidean space.

$$\omega = \omega_1 dx^1 + \omega_2 dx^2 + \omega_3 dx^3$$

Then

$$\star\omega = \omega_1\, dx^2 \wedge dx^3 - \omega_2\, dx^1 \wedge dx^3 + \omega_3\, dx^1 \wedge dx^2$$

and hence

$$\begin{aligned} d\star\omega &= \frac{\partial\omega_1}{\partial x^1}\, dx^1 \wedge dx^2 \wedge dx^3 - \frac{\partial\omega_2}{\partial x^2}\, dx^2 \wedge dx^1 \wedge dx^3 + \frac{\partial\omega_3}{\partial x^3}\, dx^3 \wedge dx^1 \wedge dx^2 \\ &= \left(\frac{\partial\omega_1}{\partial x^1} + \frac{\partial\omega_2}{\partial x^2} + \frac{\partial\omega_3}{\partial x^3}\right) dx^1 \wedge dx^2 \wedge dx^3 = (\operatorname{div}\omega)\, dx^1 \wedge dx^2 \wedge dx^3 \end{aligned} \tag{7.43}$$

We get a scalar out of this if we take the $\star$ operator again and if we use

$$\star\, dx^1 \wedge dx^2 \wedge dx^3 = 1$$

To justify this expression we define the Hodge star of a 0-form (a function) f to be its multiplication with the volume form Ω, i.e. we define

$$\star f = f\,\Omega \qquad \star\, f\,\Omega = f$$

With this definition we find

$$\star\, d\star\omega = \frac{\partial\omega_1}{\partial x^1} + \frac{\partial\omega_2}{\partial x^2} + \frac{\partial\omega_3}{\partial x^3} = \operatorname{div}\omega$$

which gives the divergence of the vector $(\omega_1, \omega_2, \omega_3)$.

From these examples we see that the standard div, grad, curl operators correspond to

$$\begin{aligned} \text{Gradient} &\quad\leftrightarrow\quad d \\ \text{Curl} &\quad\leftrightarrow\quad \star\, d \\ \text{Divergence} &\quad\leftrightarrow\quad \star\, d\,\star \end{aligned} \tag{7.44}$$

With this realization we express these vector operations in an elegant way in general coordinates, for general metrics and dimensions.
We can use the properties of the operators in Eq.(7.44) to find the forms of div, grad and curl in general orthogonal coordinates systems in three dimensions, i.e. coordinate systems for which the metric is of the form

$$g = \lambda_1^2\, dx^1 \otimes dx^1 + \lambda_2^2\, dx^2 \otimes dx^2 + \lambda_3^2\, dx^3 \otimes dx^3$$

For instance, in spherical coordinates (r, θ, ϕ) we have

$$g = dr \otimes dr + r^2 d\theta \otimes d\theta + r^2 \sin\theta d\phi \otimes d\phi$$

and hence

$$(\lambda_1, \lambda_2, \lambda_3) = (1, r, \sin\theta)$$

We can then define an orthonormal dual basis

$$e^j = \lambda_j\, dx^j \tag{7.45}$$

such that
$$g = e^1 \otimes e^1 + e^2 \otimes e^2 + e^3 \otimes e^3$$
The basis in Eq. (7.45) is dual to the vector basis
$$e_j = \frac{1}{\lambda_j}\frac{\partial}{\partial x^j}$$
such that $g(e_i, e_j) = \delta_{ij}$. In the case of spherical coordinates we have
$$e_1 = \frac{\partial}{\partial r} \quad , \quad e_2 = \frac{1}{r}\frac{\partial}{\partial \theta} \quad , \quad e_3 = \frac{1}{r\sin\theta}\frac{\partial}{\partial \phi}$$
The gradient of a function in an orthonormal basis is given by
$$df = \sum_{j=1}^{3} \frac{\partial f}{\partial x^j}\, dx^j = \sum_{j=1}^{3} \frac{1}{\lambda_j}\frac{\partial f}{\partial x^j}\, e^j$$
or
$$\nabla f = \left(\frac{1}{\lambda_1}\frac{\partial f}{\partial x^1}, \frac{1}{\lambda_2}\frac{\partial f}{\partial x^2}, \frac{1}{\lambda_3}\frac{\partial f}{\partial x^3}\right) \tag{7.46}$$
In the case of spherical coordinates we have
$$\nabla f = \left(\frac{\partial f}{\partial r}, \frac{1}{r}\frac{\partial f}{\partial \theta}, \frac{1}{r\sin\theta}\frac{\partial f}{\partial \phi}\right)$$
Let us continue with the curl for which we need to calculate $\star d$. We have for the covector ω in an orthonormal basis
$$\omega = \omega_1\, e^1 + \omega_2\, e^2 + \omega_3\, e^3 = \lambda_1\omega_1\, dx^1 + \lambda_2\omega_2\, dx^2 + \lambda_3\omega_3\, dx^3 \tag{7.47}$$
Then
$$\begin{aligned} d\omega = &\left(\frac{\partial(\lambda_2\omega_2)}{\partial x^1} - \frac{\partial(\lambda_1\omega_1)}{\partial x^2}\right) dx^1 \wedge dx^2 + \left(\frac{\partial(\lambda_3\omega_3)}{\partial x^1} - \frac{\partial(\lambda_1\omega_1)}{\partial x^3}\right) dx^1 \wedge dx^3 \\ &+ \left(\frac{\partial(\lambda_3\omega_3)}{\partial x^2} - \frac{\partial(\lambda_2\omega_2)}{\partial x^3}\right) dx^2 \wedge dx^3 \\ =&\frac{1}{\lambda_1\lambda_2}\left(\frac{\partial(\lambda_2\omega_2)}{\partial x^1} - \frac{\partial(\lambda_1\omega_1)}{\partial x^2}\right) e^1 \wedge e^2 + \frac{1}{\lambda_1\lambda_3}\left(\frac{\partial(\lambda_3\omega_3)}{\partial x^1} - \frac{\partial(\lambda_1\omega_1)}{\partial x^3}\right) e^1 \wedge e^3 \\ &+ \frac{1}{\lambda_2\lambda_3}\left(\frac{\partial(\lambda_3\omega_3)}{\partial x^2} - \frac{\partial(\lambda_2\omega_2)}{\partial x^3}\right) e^2 \wedge e^3 \end{aligned}$$
We now only need to take the $\star$ operator of this expression which gives
$$\star\, d\omega = (\nabla \times \omega)_1\, e^1 + (\nabla \times \omega)_2\, e^2 + (\nabla \times \omega)_3\, e^3$$
where
$$\nabla \times \omega = \begin{pmatrix} \frac{1}{\lambda_2\lambda_3}\left(\frac{\partial(\lambda_3\omega_3)}{\partial x^2} - \frac{\partial(\lambda_2\omega_2)}{\partial x^3}\right) \\ \frac{1}{\lambda_1\lambda_3}\left(\frac{\partial(\lambda_1\omega_1)}{\partial x^3} - \frac{\partial(\lambda_3\omega_3)}{\partial x^1}\right) \\ \frac{1}{\lambda_1\lambda_2}\left(\frac{\partial(\lambda_2\omega_2)}{\partial x^1} - \frac{\partial(\lambda_1\omega_1)}{\partial x^2}\right) \end{pmatrix} \tag{7.48}$$

In the case of spherical coordinates this gives

$$\nabla \times \omega = \begin{pmatrix} \frac{1}{r^2 \sin\theta}\left(\frac{\partial(r\sin\theta\omega_3)}{\partial\theta} - \frac{\partial(r\omega_2)}{\partial\phi}\right) \\ \frac{1}{r\sin\theta}\left(\frac{\partial(\omega_1)}{\partial\phi} - \frac{\partial(r\sin\theta\omega_3)}{\partial r}\right) \\ \frac{1}{r}\left(\frac{\partial(r\omega_2)}{\partial r} - \frac{\partial(\omega_1)}{\partial\theta}\right) \end{pmatrix} = \begin{pmatrix} \frac{1}{r\sin\theta}\left(\frac{\partial(\sin\theta\omega_3)}{\partial\theta} - \frac{\partial\omega_2}{\partial\phi}\right) \\ \frac{1}{r\sin\theta}\frac{\partial\omega_1}{\partial\phi} - \frac{1}{r}\frac{\partial(r\omega_3)}{\partial r} \\ \frac{1}{r}\left(\frac{\partial(r\omega_2)}{\partial r} - \frac{\partial\omega_1}{\partial\theta}\right) \end{pmatrix}$$

Finally we consider the divergence. Let take again a covector of the form of Eq.(7.47). Then

$$\begin{aligned} \star\omega &= \omega_1\, e^2 \wedge e^3 + \omega_2\, e^3 \wedge e^1 + \omega_3\, e^1 \wedge e^2 \\ &= \omega_1\lambda_2\lambda_3\, dx^2 \wedge dx^3 + \omega_2\lambda_1\lambda_3\, dx^3 \wedge dx^1 + \omega_3\lambda_1\lambda_2\, dx^1 \wedge dx^2 \end{aligned}$$

Then

$$\begin{aligned} d\star\omega &= \left[\frac{\partial}{\partial x^1}(\omega_1\lambda_2\lambda_3) + \frac{\partial}{\partial x^2}(\omega_2\lambda_1\lambda_3) + \frac{\partial}{\partial x^3}(\omega_3\lambda_1\lambda_2)\right] dx^1 \wedge dx^2 \wedge dx^3 \\ &= \frac{1}{\lambda_1\lambda_2\lambda_3}\left[\frac{\partial}{\partial x^1}(\omega_1\lambda_2\lambda_3) + \frac{\partial}{\partial x^2}(\omega_2\lambda_1\lambda_3) + \frac{\partial}{\partial x^3}(\omega_3\lambda_1\lambda_2)\right] e^1 \wedge e^2 \wedge e^3 \end{aligned}$$

and therefore

$$\star d\star\omega = \frac{1}{\lambda_1\lambda_2\lambda_3}\left[\frac{\partial}{\partial x^1}(\omega_1\lambda_2\lambda_3) + \frac{\partial}{\partial x^2}(\omega_2\lambda_1\lambda_3) + \frac{\partial}{\partial x^3}(\omega_3\lambda_1\lambda_2)\right] = \operatorname{div}\omega \tag{7.49}$$

In the case of spherical coordinates this becomes

$$\begin{aligned} \operatorname{div}\omega &= \frac{1}{r^2\sin\theta}\left[\frac{\partial}{\partial r}(r^2\sin\theta\omega_1) + \frac{\partial}{\partial\theta}(r\sin\theta\omega_2) + \frac{\partial}{\partial\phi}(r\omega_3)\right] \\ &= \frac{1}{r^2}\frac{\partial}{\partial r}(r^2\omega_1) + \frac{1}{r\sin\theta}\frac{\partial}{\partial\theta}(\sin\theta\omega_2) + \frac{1}{r\sin\theta}\frac{\partial}{\partial\phi}\omega_3 \end{aligned}$$

We have covered most familiar vector operations. The one that we are missing is the Laplacian. But since

$$\nabla^2 f = \operatorname{div}(\operatorname{grad} f) = \star d\star df \tag{7.50}$$

we only need to insert the components of Eq.(7.46) for ω_i into Eq.(7.49) which yields

$$\nabla^2 f = \frac{1}{\lambda_1\lambda_2\lambda_3}\left[\frac{\partial}{\partial x^1}(\frac{\lambda_2\lambda_3}{\lambda_1}\frac{\partial f}{\partial x^1}) + \frac{\partial}{\partial x^1}(\frac{\lambda_1\lambda_3}{\lambda_2}\frac{\partial f}{\partial x^2}) + \frac{\partial}{\partial x^1}(\frac{\lambda_1\lambda_2}{\lambda_3}\frac{\partial f}{\partial x^3})\right]$$

Again for the spherical coordinates this gives

$$\begin{aligned} \nabla^2 f &= \frac{1}{r^2\sin\theta}\left[\frac{\partial}{\partial r}\left(r^2\sin\theta\frac{\partial f}{\partial r}\right) + \frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial f}{\partial\theta}\right) + \frac{\partial}{\partial\phi}\left(\frac{1}{\sin\theta}\frac{\partial f}{\partial\phi}\right)\right] \\ &= \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial f}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial f}{\partial\theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2 f}{\partial\phi^2} \end{aligned}$$

7.5 Maxwell's equations from exterior calculus

From the examples of the previous section we may expect that Maxwell's equations can de derived by taking the exterior derivative of the electromagnetic field tensor. This is indeed the case. More precisely the four Maxwell equations are given by

$$dF = 0 \tag{7.51}$$

$$\star d\star F = \frac{4\pi}{c}\,J \tag{7.52}$$

where J is the current one-form

$$J = \sum_{\mu=0}^{3} J_\mu dx^\mu \tag{7.53}$$

obtained from the current four-vector

$$j = \sum_{\mu=0}^{3} j^\mu \frac{\partial}{\partial x^\mu}$$

by lowering the indices,

$$J_\mu = \sum_{\nu=0}^{3} g_{\mu\nu} j^\nu \tag{7.54}$$

Let us check Eqs.(7.51) and (7.52). We write

$$\begin{aligned} F = \sum_{\mu,\nu=0}^{3} F_{\mu\nu}\, dx^\mu \otimes dx^\nu &= \sum_{\mu<\nu=0}^{3} F_{\mu\nu}\, dx^\mu \wedge dx^\nu \\ &= - E_1\, dx^0 \wedge dx^1 - E_2\, dx^0 \wedge dx^2 - E_3\, dx^0 \wedge dx^3 \\ &\quad + B_3\, dx^1 \wedge dx^2 - B_2\, dx^1 \wedge dx^3 + B_1\, dx^2 \wedge dx^3 \end{aligned} \tag{7.55}$$

We have

$$\begin{aligned} dF =& - \frac{\partial E_1}{\partial x^2} dx^2 \wedge dx^0 \wedge dx^1 - \frac{\partial E_1}{\partial x^3} dx^3 \wedge dx^0 \wedge dx^1 - \frac{\partial E_2}{\partial x^1} dx^1 \wedge dx^0 \wedge dx^2 \\ & - \frac{\partial E_2}{\partial x^3} dx^3 \wedge dx^0 \wedge dx^2 - \frac{\partial E_3}{\partial x^1} dx^1 \wedge dx^0 \wedge dx^3 - \frac{\partial E_3}{\partial x^2} dx^2 \wedge dx^0 \wedge dx^3 \\ & + \frac{\partial B_3}{\partial x^0} dx^0 \wedge dx^1 \wedge dx^2 + \frac{\partial B_3}{\partial x^3} dx^3 \wedge dx^1 \wedge dx^2 - \frac{\partial B_2}{\partial x^0} dx^0 \wedge dx^1 \wedge dx^3 \\ & - \frac{\partial B_2}{\partial x^2} dx^2 \wedge dx^1 \wedge dx^3 + \frac{\partial B_1}{\partial x^0} dx^0 \wedge dx^2 \wedge dx^3 + \frac{\partial B_1}{\partial x^1} dx^1 \wedge dx^2 \wedge dx^3 \\ =& \left[\frac{\partial E_2}{\partial x^1} - \frac{\partial E_1}{\partial x^2} + \frac{\partial B_3}{\partial x^0}\right] dx^0 \wedge dx^1 \wedge dx^2 + \left[\frac{\partial E_1}{\partial x^3} - \frac{\partial E_3}{\partial x^1} + \frac{\partial B_2}{\partial x^0}\right] dx^0 \wedge dx^3 \wedge dx^1 \\ & + \left[\frac{\partial E_3}{\partial x^2} - \frac{\partial E_2}{\partial x^3} + \frac{\partial B_1}{\partial x^0}\right] dx^0 \wedge dx^2 \wedge dx^3 + \left[\frac{\partial B_1}{\partial x^1} + \frac{\partial B_2}{\partial x^2} + \frac{\partial B_3}{\partial x^3}\right] dx^1 \wedge dx^2 \wedge dx^3 \end{aligned}$$

We therefore see that if we require $dF = 0$ we have

$$dF = 0 \quad \leftrightarrow \quad \nabla \cdot \mathbf{B} = 0 \quad , \quad \nabla \times \mathbf{E} + \frac{1}{c}\frac{\partial \mathbf{B}}{\partial t} = 0 \tag{7.56}$$

which is equivalent to the Maxwell equations (7.22) and (7.23). Let us now evaluate $\star d \star F$. We assume that we have a standard Minkowskian metric. We then have (see Eq.(3.104))

$$\begin{aligned} \star F =& B_1\, dx^0 \wedge dx^1 + B_2\, dx^0 \wedge dx^2 + B_3\, dx^0 \wedge dx^3 \\ & + E_3\, dx^1 \wedge dx^2 - E_2\, dx^1 \wedge dx^3 + E_1\, dx^2 \wedge dx^3 \end{aligned}$$

This expression is obtained simply by the replacement $\mathbf{E} \to -\mathbf{B}$ and $\mathbf{E} \to \mathbf{B}$ in F. By making a similar replacement in dF we find therefore immediately that

$$\begin{aligned} d \star F =& \left[-\frac{\partial B_2}{\partial x^1} + \frac{\partial B_1}{\partial x^2} + \frac{\partial E_3}{\partial x^0}\right] dx^0 \wedge dx^1 \wedge dx^2 + \left[-\frac{\partial B_1}{\partial x^3} + \frac{\partial B_3}{\partial x^1} + \frac{\partial E_2}{\partial x^0}\right] dx^0 \wedge dx^3 \wedge dx^1 \\ & + \left[-\frac{\partial B_3}{\partial x^2} + \frac{\partial B_2}{\partial x^3} + \frac{\partial E_1}{\partial x^0}\right] dx^0 \wedge dx^2 \wedge dx^3 + \left[\frac{\partial E_1}{\partial x^1} + \frac{\partial E_2}{\partial x^2} + \frac{\partial E_3}{\partial x^3}\right] dx^1 \wedge dx^2 \wedge dx^3 \end{aligned}$$

We finally need to act with the $\star$ operator on this expression. We have using Eq.(3.95) that

$$\begin{aligned}
\star\, dx^0\wedge dx^1\wedge dx^2 &= \epsilon_{0123}\, g^{00}g^{11}g^{22}\, dx^3 = -dx^3\\
\star\, dx^0\wedge dx^3\wedge dx^1 &= \epsilon_{0312}\, g^{00}g^{11}g^{33}\, dx^2 = -dx^2\\
\star\, dx^0\wedge dx^2\wedge dx^3 &= \epsilon_{0231}\, g^{00}g^{22}g^{33}\, dx^1 = -dx^1\\
\star\, dx^1\wedge dx^2\wedge dx^3 &= \epsilon_{1230}\, g^{11}g^{22}g^{33}\, dx^0 = -dx^0
\end{aligned}$$

and find that

$$\begin{aligned}
\star\, d\star F = &\left[-\frac{\partial E_1}{\partial x^1}-\frac{\partial E_2}{\partial x^2}-\frac{\partial E_3}{\partial x^3}\right] dx^0+\left[\frac{\partial B_3}{\partial x^2}-\frac{\partial B_2}{\partial x^3}-\frac{\partial E_1}{\partial x^0}\right] dx^1\\
&+\left[\frac{\partial B_1}{\partial x^3}-\frac{\partial B_3}{\partial x^1}-\frac{\partial E_2}{\partial x^0}\right] dx^2+\left[\frac{\partial B_2}{\partial x^1}-\frac{\partial B_1}{\partial x^2}-\frac{\partial E_3}{\partial x^0}\right] dx^3
\end{aligned}\tag{7.57}$$

Furthermore the current one-form J is given by

$$J = J_0\, dx^0 + J_1\, dx^1 + J_2\, dx^2 + J_3\, dx^3 = -c\rho\, dx^0 + j^1\, dx^1 + j^2\, dx^2 + j^3\, dx^3$$

where we used Eq.(7.54) and $j = (c\rho, \mathbf{j})$. Equating this equation with (7.57) we see that

$$\star\, d\star F = \frac{4\pi}{c}J \quad \leftrightarrow \quad \nabla\cdot\mathbf{E} = 4\pi\rho \quad , \quad \nabla\times\mathbf{B} - \frac{1}{c}\frac{\partial\mathbf{E}}{\partial t} = \frac{4\pi}{c}\mathbf{j} \tag{7.58}$$

We therefore recover the Maxwell equations (7.20) and (7.21). Our analysis therefore shows that Eqs.(7.51) and (7.52) indeed comprise the full set of the four Maxwell equations.

7.6 Comparing two formulations

We have thus found two ways to express Maxwell's equations. First of all we have Eqs.(7.25) and (7.26) in terms of the covariant derivative, and secondly we have Eqs.(7.51) and (7.52) in terms of the exterior derivative. These formulations are clearly related, so let us now establish this final link. We may expect that there is such a link since we have seen that $\star d\star$ is related to a divergence and Eqs. (7.25) and (7.26) are given as a divergence. However, we only checked this relation for vectors. So what we need to do is to check it for tensors as well. If ω is an anti-symmetric tensor of order p (a p-form) on an n-dimensional space i.e.

$$\omega = \sum_{i_1<\ldots<i_p}^{n} \omega_{i_1\ldots i_p} dx^{i_1}\wedge\ldots\wedge dx^{i_p} \tag{7.59}$$

then $\star d\star\omega$ is a $(p-1)$-form

$$\star\, d\star\omega = \sum_{m_1<\ldots<m_{p-1}}^{n} (\star\, d\star\omega)_{m_1\ldots m_{p-1}}\, dx^{m_1}\wedge\ldots\wedge dx^{m_{p-1}} \tag{7.60}$$

with coefficients given by

$$\begin{aligned}
&(\star\, d\star\omega)_{m_1\ldots m_{p-1}} =\\
&\quad\frac{\mathrm{sign}(g)}{\sqrt{|g|}}(-1)^{np+n}\sum_{k,r_1\ldots r_{p-1}}^{n}\frac{\partial}{\partial x^k}(\sqrt{|g|}\,\omega^{kr_1\ldots r_{p-1}})g_{r_1m_1}\ldots g_{r_{p-1}m_{p-1}}
\end{aligned}\tag{7.61}$$

The proof of this expression is a straightforward, albeit tedious, derivation using the definitions of the d and the $\star$ operators. In order to no interrupt the presentation too much this derivation is given in Appendix B. This expression looks somewhat simpler if we raise the indices on $\star d \star \omega$. We have

$$(\star d \star \omega)^{r_1 \dots r_{p-1}} = \mathrm{sign}(g)(-1)^{np+n} \frac{(-1)^{p-1}}{\sqrt{|g|}} \sum_k^n \frac{\partial}{\partial x^k} (\sqrt{|g|}\, \omega^{r_1 \dots r_{p-1} k})$$

where we also moved the index k over $(p-1)$ positions to the end which gives an additional pre-factor $(-1)^{p-1}$ due to the anti-symmetry of ω. If we compare now to Eq.(7.34) we recognize the divergence of the tensor ω. We can therefore write

$$(\star d \star \omega)^{r_1 \dots r_{p-1}} = \mathrm{sign}(g)(-1)^{np+n+p+1} \sum_k^n \omega^{r_1 \dots r_{p-1} k}_{;k} \tag{7.62}$$

This is the equation that we need. We take now $\omega = F$ to be the electromagnetic field tensor. Then we have $n = 4$, $p = 2$ and $\mathrm{sign}(g) = -1$. Then Eq.(7.62) gives

$$(\star d \star F)^\mu = \sum_{\nu=0}^3 F^{\mu\nu}_{;\nu} \tag{7.63}$$

This provides the desired connection between Eqs.(7.25) and (7.52). In other words, Eq.(7.25) could also have been written as

$$(\star d \star F)^\sharp = \frac{4\pi}{c}\, j$$

which by lowering the index leads to Eq.(7.52). What about the other two Maxwell equations? Let us again consider Eq.(7.62) but now choose $\omega = \star F$. Then since according to Eq.(3.108) we have $\star \star F = -F$ it follows that

$$-(\star dF)^\mu = (\star d \star \star F)^\mu = \sum_{\nu=0}^3 (\star F)^{\mu\nu}_{;\nu} \tag{7.64}$$

Therefore Eq.(7.26) is equivalent to $\star dF = 0$ which is equivalent to $dF = 0$ and gives back Eq.(7.51). This completely establishes the link between the two formulations.

7.7 Maxwell's equations in general orthogonal curvilinear coordinates

Many of the cases in which we can solve the Maxwell equations analytically involve a special symmetry. The equations simplify if we use a symmetry-adapted coordinate system, such as cylindrical or spherical coordinates. It is therefore worthwhile to investigate the form of the Maxwell equations in such coordinate systems. Let the original metric be

$$g = -dx^0 \otimes dx^0 + dx^1 \otimes dx^1 + dx^2 \otimes dx^2 + dx^3 \otimes dx^3$$

corresponding to a standrad Minkowskian metric. We now assume that the problem has a certain spatial symmetry and we therefore want to transform the spatial coordinates. We then have $x^i(y^1, y^2, y^3)$ for $i = 1, 2, 3$ and $x^0 = y^0 = ct$. If we define

$$\frac{\partial \mathbf{x}}{\partial y^i} = \left(\frac{\partial x^1}{\partial y^i}, \frac{\partial x^2}{\partial y^i}, \frac{\partial x^3}{\partial y^i} \right)$$

then we say that the new coordinate system is orthogonal when

$$g_{ij} = \langle \frac{\partial \mathbf{x}}{\partial y^i}, \frac{\partial \mathbf{x}}{\partial y^j} \rangle = \lambda_i^2 \, \delta_{ij} \tag{7.65}$$

This means that the metric in coordinate system (y^0, y^1, y^2, y^3) attains the form

$$g = -dy^0 \otimes dy^0 + \lambda_1^2 \, dy^1 \otimes dy^1 + \lambda_2^2 \, dy^2 \otimes dy^2 + \lambda_3^2 \, dy^3 \otimes dy^3$$

Let us now see how the electromagnetic field tensor transforms to this coordinate system. We define the electric field one-form

$$E = E_1 \, dx^1 + E_2 \, dx^2 + E_3 \, dx^3 \tag{7.66}$$

and the magnetic field two-form

$$B = B_1 \, dx^2 \wedge dx^3 + B_2 \, dx^3 \wedge dx^1 + B_3 \, dx^1 \wedge dx^2 \tag{7.67}$$

in terms of the original coordinate system. From Eq.(7.55) we then see that the electromagnetic field tensor can be written as

$$F = B + E \wedge dx^0$$

Now the transformation law for the electric field is quite straightforward. We have

$$E = \sum_{j=1}^{3} E_j \, dx^j = \sum_{j,k=1}^{3} E_j \frac{\partial x^j}{\partial y^k} dy^k = \sum_{k=1}^{3} E'_k \, dy^k$$

So we have

$$\mathbf{E}' = \left(\frac{\partial \mathbf{x}}{\partial y^1}, \frac{\partial \mathbf{x}}{\partial y^2}, \frac{\partial \mathbf{x}}{\partial y^3} \right)^T \mathbf{E} = L \, \mathbf{E} \tag{7.68}$$

where L is a matrix with as rows the vectors $\partial \mathbf{x}/\partial y^i$. For the magnetic field two-form we have

$$\begin{aligned}
B &= \sum_{j,k=1}^{3} \left[B_1 \frac{\partial x^2}{\partial y^j} \frac{\partial x^3}{\partial y^k} + B_2 \frac{\partial x^3}{\partial y^j} \frac{\partial x^1}{\partial y^k} + B_3 \frac{\partial x^1}{\partial y^j} \frac{\partial x^2}{\partial y^k} \right] dy^j \wedge dy^k \\
&= \sum_{j<k}^{3} \left[B_1 \left(\frac{\partial x^2}{\partial y^j} \frac{\partial x^3}{\partial y^k} - \frac{\partial x^3}{\partial y^j} \frac{\partial x^2}{\partial y^k} \right) + B_2 \left(\frac{\partial x^3}{\partial y^j} \frac{\partial x^1}{\partial y^k} - \frac{\partial x^1}{\partial y^j} \frac{\partial x^3}{\partial y^k} \right) \right. \\
&\quad \left. + B_3 \left(\frac{\partial x^1}{\partial y^j} \frac{\partial x^2}{\partial y^k} - \frac{\partial x^2}{\partial y^j} \frac{\partial x^1}{\partial y^k} \right) \right] dy^j \wedge dy^k \\
&= B'_1 \, dy^2 \wedge dy^3 + B'_2 \, dy^3 \wedge dy^1 + B'_3 \, dy^1 \wedge dy^2
\end{aligned}$$

from which we see that

$$\mathbf{B}' = \left(\frac{\partial \mathbf{x}}{\partial y^2} \times \frac{\partial \mathbf{x}}{\partial y^3}, \frac{\partial \mathbf{x}}{\partial y^3} \times \frac{\partial \mathbf{x}}{\partial y^1}, \frac{\partial \mathbf{x}}{\partial y^1} \times \frac{\partial \mathbf{x}}{\partial y^2} \right)^T \mathbf{B} = M \, \mathbf{B} \tag{7.69}$$

where M is a matrix that contains as rows the outer products of the vectors $\partial \mathbf{x}/\partial y^j$. So far our derivations apply to general spatial coordinate transformations. However, if we use an orthogonal coordinate system for which Eq.(7.65) holds then the outer product of $\partial \mathbf{x}/\partial y^i$ and $\partial \mathbf{x}/\partial y^j$ is proportional to $\partial \mathbf{x}/\partial y^k$ for $i \neq j \neq k$. For instance

$$\frac{\partial \mathbf{x}}{\partial y^2} \times \frac{\partial \mathbf{x}}{\partial y^3} = \alpha \, \frac{\partial \mathbf{x}}{\partial y^1}$$

where α is a factor to be determined. Since these three vectors are orthogonal we have

$$\lambda_2\lambda_3 = \left|\frac{\partial \mathbf{x}}{\partial y^2} \times \frac{\partial \mathbf{x}}{\partial y^3}\right| = |\alpha| \left|\frac{\partial \mathbf{x}}{\partial y^1}\right| = \lambda_1 \, |\alpha|$$

and hence

$$|\alpha| = \frac{\lambda_2\lambda_3}{\lambda_1}$$

This determines the value of α up to a sign. If we, however, impose that our coordinate transformation is orientation preserving, meaning that the Jacobian of the transformation is positive

$$J = \det(\frac{\partial \mathbf{x}}{\partial y^1}, \frac{\partial \mathbf{x}}{\partial y^2}, \frac{\partial \mathbf{x}}{\partial y^3}) > 0 \tag{7.70}$$

then it follows from

$$\det(\mathbf{abc}) = \mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = \mathbf{b} \cdot (\mathbf{c} \times \mathbf{a}) = \mathbf{c} \cdot (\mathbf{a} \times \mathbf{b})$$

that $\alpha >$ for all the three outer products in Eq.(7.69). It then follows that the transformation is given by

$$\mathbf{B}' = \left(\frac{\lambda_2\lambda_3}{\lambda_1}\frac{\partial \mathbf{x}}{\partial y^1}, \frac{\lambda_1\lambda_3}{\lambda_2}\frac{\partial \mathbf{x}}{\partial y^2}, \frac{\lambda_1\lambda_2}{\lambda_3}\frac{\partial \mathbf{x}}{\partial y^3}\right)^T \mathbf{B} \tag{7.71}$$

From this expression we see that it will be advantageous to work in an orthonormal basis. If we define

$$e^i = \lambda_i \, dy^i \tag{7.72}$$

as well as $e^0 = dy^0$ then the metric g attains the form

$$g = -e^0 \otimes e^0 + e^1 \otimes e^1 + e^2 \otimes e^2 + e^3 \otimes e^3$$

We can now wrtite the electric field one-form and the magnetic field two-form in this basis as

$$\begin{aligned} E &= E_1'' \, e^1 + E_2'' \, e^2 + E_3'' \, e^3 \\ B &= B_1'' \, e^2 \wedge e^3 + B_2'' \, e^3 \wedge e^1 + B_3'' \, e^1 \wedge e^2 \end{aligned}$$

where

$$E_j'' = \frac{E_j'}{\lambda_j}$$

and

$$B_1'' = \frac{B_1'}{\lambda_2\lambda_3} \quad , \quad B_2'' = \frac{B_2'}{\lambda_1\lambda_3} \quad , \quad B_3'' = \frac{B_3'}{\lambda_1\lambda_2}$$

In this new orthonormal basis the transformation laws (7.68) and (7.69) become

$$\begin{aligned} \mathbf{E}'' &= \left(\frac{1}{\lambda_1}\frac{\partial \mathbf{x}}{\partial y^1}, \frac{1}{\lambda_2}\frac{\partial \mathbf{x}}{\partial y^2}, \frac{1}{\lambda_3}\frac{\partial \mathbf{x}}{\partial y^3}\right)^T \mathbf{E} = N\mathbf{E} \\ \mathbf{B}'' &= \left(\frac{1}{\lambda_1}\frac{\partial \mathbf{x}}{\partial y^1}, \frac{1}{\lambda_2}\frac{\partial \mathbf{x}}{\partial y^2}, \frac{1}{\lambda_3}\frac{\partial \mathbf{x}}{\partial y^3}\right)^T \mathbf{B} = N\mathbf{B} \end{aligned}$$

where now the row vectors $(1/\lambda_j)\partial\mathbf{x}/\partial y^j$ of the matrix N form an orthonormal set. So with this orthonormal basis it appears that the electric and magnetic fields transform in a similar way under spatial transformations. This, however, only appearance. If we do not use orthogonal coordinates, or use transformations that change the orientation then they transform differently. For instance, if we perform a space inversion

$$x^i \to -x^i = y^i$$

in Eqs.(7.66) and (7.67) then we see that

$$E'_i = -E_i \quad , \quad B'_i = B_i$$

such that the electric field transforms to minus itself but the magnetic field remains the same. For this reason it is sometimes said that the $\mathbf{B}$-field is a pseudo-vector. This is, however, somewhat confusing. The most precise things to say is that the $\mathbf{B}$-field is a two-form.
As a nice application of the Hodge $\star$ operator let us show that the transformation law for the magnetic field could have been derived differently as well. From Eq.(7.67) it follows that

$$\star B = B_1\, dx^1 + B_2 dx^2 + B_3 dx^3$$

Transforming this to coordinate system y gives

$$\star B = \sum_{i,j=1}^{3} B_i \frac{\partial x^i}{\partial y^j} dy^j \tag{7.73}$$

If we, however, first write B in new coordinates we have

$$B = B'_1\, dy^2 \wedge dy^3 + B'_2\, dy^3 \wedge dy^1 + B'_3\, dy^1 \wedge dy^2$$

We can now take the star operation on this expression using Eq.(3.97). This gives

$$\star\, dy^2 \wedge dy^3 = \epsilon_{231} \sqrt{|g|} \begin{vmatrix} g^{22} & g^{23} \\ g^{32} & g^{33} \end{vmatrix} dy^1 = \lambda_1 \lambda_2 \lambda_2 \begin{vmatrix} \frac{1}{\lambda_2^2} & 0 \\ 0 & \frac{1}{\lambda_3^2} \end{vmatrix} dy^1 = \frac{\lambda_1}{\lambda_2 \lambda_3} dy^1$$

and similarly

$$\star\, dy^3 \wedge dy^1 = \frac{\lambda_2}{\lambda_1 \lambda_3} dy^2 \quad , \quad \star\, dy^1 \wedge dy^2 = \frac{\lambda_3}{\lambda_1 \lambda_2} dy^3$$

We therefore find that

$$\star B = \frac{\lambda_1 B'_1}{\lambda_2 \lambda_3}\, dy^1 + \frac{\lambda_2 B'_2}{\lambda_1 \lambda_3}\, dy^2 + \frac{\lambda_3 B'_3}{\lambda_1 \lambda_2}\, dy^3$$

Comparing this expression to Eq.(7.73) then gives

$$\frac{\lambda_1 B'_1}{\lambda_2 \lambda_3} = \sum_{i=1}^{3} B_i \frac{\partial x^i}{\partial y^1} \qquad \frac{\lambda_2 B'_2}{\lambda_1 \lambda_3} = \sum_{i=1}^{3} B_i \frac{\partial x^i}{\partial y^2} \qquad \frac{\lambda_3 B'_3}{\lambda_1 \lambda_2} = \sum_{i=1}^{3} B_i \frac{\partial x^i}{\partial y^3}$$

We see that we recovered exactly the transformation law of Eq.(7.71).
Let us now look at Maxwell's equations again. We saw that it is convenient to work in an orthonormal basis. The basis e_j dual to the basis in Eq.(7.72) is given by

$$e_j = \frac{1}{\lambda_j} \frac{\partial}{\partial y^j} \qquad j = 1, 2, 3$$

and $e_0 = \partial / \partial y^0$. The current four-vector in this basis is given by

$$j = j^0 e_0 + j^1 e_1 + j^2 e_2 + j^3 e_3 = c\rho \frac{\partial}{\partial y^0} + \frac{j^1}{\lambda_1} \frac{\partial}{\partial y^1} + \frac{j^1}{\lambda_2} \frac{\partial}{\partial y^2} + \frac{j^1}{\lambda_3} \frac{\partial}{\partial y^3} \tag{7.74}$$

In terms of the dual vectors e_j the electromagnetic field tensor is given by

$$F = -E_1\, e^0 \wedge e^1 - E_2\, e^0 \wedge e^2 - E_3\, e^0 \wedge e^3 + B_1\, e^1 \wedge e^2 + B_2\, e^3 \wedge e^1 + B_3\, e^1 \wedge e^2 \tag{7.75}$$

whereas is contravariant version with components $F^{\mu\nu}$, which we will denote by $F^\sharp$ is given by

$$F^\sharp = E_1\, e_0 \wedge e_1 + E_2\, e_0 \wedge e_2 + E_3\, e_0 \wedge e_3 + B_1\, e_1 \wedge e_2 + B_2\, e_3 \wedge e_1 + B_3\, e_1 \wedge e_2$$

Let us now see what equations we obtain from the Maxwell Eqs.(7.25) or equivalently Eq.(7.33). To apply Eq.(7.33) we first write $F^\sharp$ in coordinate basis as

$$\begin{aligned} F^\sharp =& \frac{E_1}{\lambda_1}\frac{\partial}{\partial y^0} \wedge \frac{\partial}{\partial y^1} + \frac{E_2}{\lambda_2}\frac{\partial}{\partial y^0} \wedge \frac{\partial}{\partial y^2} + \frac{E_3}{\lambda_3}\frac{\partial}{\partial y^0} \wedge \frac{\partial}{\partial y^3} \\ &+ \frac{B_1}{\lambda_2\lambda_3}\frac{\partial}{\partial y^2} \wedge \frac{\partial}{\partial y^3} + \frac{B_2}{\lambda_1\lambda_3}\frac{\partial}{\partial y^3} \wedge \frac{\partial}{\partial y^1} + \frac{B_3}{\lambda_1\lambda_2}\frac{\partial}{\partial y^1} \wedge \frac{\partial}{\partial y^2} \end{aligned}$$

and hence

$$F^{\mu\nu} = \begin{pmatrix} 0 & \frac{E_1}{\lambda_1} & \frac{E_2}{\lambda_2} & \frac{E_3}{\lambda_3} \\ -\frac{E_1}{\lambda_1} & 0 & \frac{B_3}{\lambda_1\lambda_2} & -\frac{B_2}{\lambda_1\lambda_3} \\ -\frac{E_2}{\lambda_2} & -\frac{B_3}{\lambda_1\lambda_2} & 0 & \frac{B_1}{\lambda_2\lambda_3} \\ -\frac{E_3}{\lambda_3} & \frac{B_2}{\lambda_1\lambda_3} & -\frac{B_1}{\lambda_2\lambda_3} & 0 \end{pmatrix}$$

We further have

$$\sqrt{|g|} = \sqrt{|g_{00}g_{11}g_{22}g_{33}|} = \lambda_1\lambda_2\lambda_3$$

Then from Eq.(7.33) and Eq.(7.74) it follows that

$$\begin{aligned} 4\pi\rho &= \frac{1}{\lambda_1\lambda_2\lambda_3}\left[\frac{\partial}{\partial y^1}(\lambda_2\lambda_3 E_1) + \frac{\partial}{\partial y^2}(\lambda_1\lambda_3 E_2) + \frac{\partial}{\partial y^3}(\lambda_1\lambda_2 E_3)\right] \\ \frac{4\pi}{c}\frac{j^1}{\lambda_1} &= \frac{1}{\lambda_1\lambda_2\lambda_3}\left[-\frac{\partial}{\partial y^0}(\lambda_2\lambda_3 E_1) + \frac{\partial}{\partial y^2}(\lambda_3 B_3) - \frac{\partial}{\partial y^3}(\lambda_2 B_2)\right] \\ \frac{4\pi}{c}\frac{j^2}{\lambda_2} &= \frac{1}{\lambda_1\lambda_2\lambda_3}\left[-\frac{\partial}{\partial y^0}(\lambda_1\lambda_3 E_2) + \frac{\partial}{\partial y^3}(\lambda_1 B_1) - \frac{\partial}{\partial y^1}(\lambda_3 B_3)\right] \\ \frac{4\pi}{c}\frac{j^1}{\lambda_1} &= \frac{1}{\lambda_1\lambda_2\lambda_3}\left[-\frac{\partial}{\partial y^0}(\lambda_1\lambda_2 E_3) + \frac{\partial}{\partial y^1}(\lambda_2 B_2) - \frac{\partial}{\partial y^2}(\lambda_1 B_1)\right] \end{aligned} \tag{7.76}$$

On the right hand side of the first equation we recognize the divergence of Eq.(7.49) in general orthogonal coordinates, whereas in the remaining three equations we recognize the formula of the curl of Eq.(7.48). Further more since the λ_i are time-independent (we only transformed the spatial coordinates) they can be taken out of the terms involving $\partial/\partial y^0$. Let us now check if we get the same from Eq.(7.52). We have

$$\begin{aligned} \star F =& B_1\, e^0 \wedge e^1 + B_2\, e^0 \wedge e^2 + B_3\, e^0 \wedge e^3 + E_1\, e^2 \wedge e^3 + E_2\, e^3 \wedge e^1 + E_3\, e^1 \wedge e^2 \\ =& B_1\lambda_1\, dy^0 \wedge dy^1 + B_1\lambda_1\, dy^0 \wedge dy^1 + B_1\lambda_1\, dy^0 \wedge dy^1 \\ &+ E_1\lambda_2\lambda_3\, dy^2 \wedge dy^3 + E_2\lambda_3\lambda_1\, dy^3 \wedge dy^1 + E_3\lambda_1\lambda_2\, dy^1 \wedge dy^2 \end{aligned}$$

The current covector is obtained by lowering the indices on j and given by

$$J = -\rho c\, e^0 + j^1\, e^1 + j^2\, e^2 + j^3\, e^3 = -c\rho\, dy^0 + j^1\lambda_1\, dy^1 + j^2\lambda_2\, dy^2 + j^3\lambda_3\, dy^3 \tag{7.77}$$

We can now take the exterior derivative of $\star F$ to obtain

$$
\begin{aligned}
d\star F =& \frac{1}{\lambda_1\lambda_2}\left[\frac{\partial}{\partial y^0}(\lambda_1\lambda_2 E_3) - \frac{\partial}{\partial y^1}(\lambda_2 B_2) + \frac{\partial}{\partial y^2}(\lambda_1 B_1)\right] e^0\wedge e^1\wedge e^2 \\
&+ \frac{1}{\lambda_1\lambda_3}\left[\frac{\partial}{\partial y^0}(\lambda_1\lambda_3 E_2) - \frac{\partial}{\partial y^3}(\lambda_1 B_1) + \frac{\partial}{\partial y^1}(\lambda_3 B_3)\right] e^0\wedge e^1\wedge e^3 \\
&+ \frac{1}{\lambda_2\lambda_3}\left[\frac{\partial}{\partial y^0}(\lambda_2\lambda_3 E_1) - \frac{\partial}{\partial y^2}(\lambda_3 B_3) + \frac{\partial}{\partial y^3}(\lambda_2 B_2)\right] e^0\wedge e^2\wedge e^3 \\
&+ \frac{1}{\lambda_1\lambda_2\lambda_3}\left[\frac{\partial}{\partial y^1}(\lambda_2\lambda_3 E_1) + \frac{\partial}{\partial y^2}(\lambda_1\lambda_3 E_2) + \frac{\partial}{\partial y^3}(\lambda_1\lambda_2 E_3)\right] e^1\wedge e^2\wedge e^3
\end{aligned}
$$

Then using

$$
\begin{aligned}
&\star e^0\wedge e^1\wedge e^2 = -e^3 \qquad && \star e^0\wedge e^3\wedge e^1 = -e^2 \\
&\star e^0\wedge e^2\wedge e^2 = -e^1 && \star e^1\wedge e^2\wedge e^3 = -e^0
\end{aligned}
$$

we find that

$$
\begin{aligned}
\star d\star F =& -\frac{1}{\lambda_1\lambda_2\lambda_3}\left[\frac{\partial}{\partial y^1}(\lambda_2\lambda_3 E_1) + \frac{\partial}{\partial y^2}(\lambda_1\lambda_3 E_2) + \frac{\partial}{\partial y^3}(\lambda_1\lambda_2 E_3)\right] e^0 \\
&+ \frac{1}{\lambda_2\lambda_3}\left[\frac{\partial}{-\partial y^0}(\lambda_2\lambda_3 E_1) + \frac{\partial}{\partial y^2}(\lambda_3 B_3) - \frac{\partial}{\partial y^3}(\lambda_2 B_2)\right] e^1 \\
&+ \frac{1}{\lambda_1\lambda_3}\left[-\frac{\partial}{\partial y^0}(\lambda_1\lambda_3 E_2) + \frac{\partial}{\partial y^3}(\lambda_1 B_1) - \frac{\partial}{\partial y^1}(\lambda_3 B_3)\right] e^2 \\
&+ \frac{1}{\lambda_1\lambda_2}\left[-\frac{\partial}{\partial y^0}(\lambda_1\lambda_2 E_3) + \frac{\partial}{\partial y^1}(\lambda_2 B_2) - \frac{\partial}{\partial y^2}(\lambda_1 B_1)\right] e^3
\end{aligned}
$$

If we compare this expression to the current covector J in Eq.(7.77) we see that the relation

$$
\star d\star F = \frac{4\pi}{c} J \tag{7.78}
$$

is indeed equivalent to the four equations (7.76). As an exercise you can yourself work out the set of equations

$$
\begin{aligned}
0 &= \frac{1}{\sqrt{|g|}}\sum_{\nu=0}^{3}\frac{\partial}{\partial x^\nu}\left(\sqrt{|g|}\,(\star F)^{\mu\nu}\right) \\
0 &= dF
\end{aligned}
$$

in the same coordinates and see that they yield the same equations.
Let us finally give the equation for the Lorentz force in the general orthogonal coordinate system. If we write out Eq.(7.75) for F in terms of a coordinate basis we have

$$
\begin{aligned}
F = & -E_1\lambda_1\, dy^0\wedge dy^1 - E_2\lambda_2\, dy^0\wedge dy^2 - E_3\lambda_3\, dy^0\wedge dy^3 \\
& + B_1\lambda_2\lambda_3\, dy^1\wedge dy^2 + B_2\lambda_1\lambda_3\, dy^3\wedge dy^1 + B_3\lambda_1\lambda_2\, dy^1\wedge dy^2
\end{aligned}
$$

and therefore

$$
F^{\mu\nu} = \begin{pmatrix}
0 & -\lambda_1 E_1 & -\lambda_2 E_2 & -\lambda_3 E_3 \\
\lambda_1 E_1 & 0 & \lambda_1\lambda_2 B_3 & -\lambda_1\lambda_3 B_2 \\
\lambda_2 E_2 & -\lambda_1\lambda_2 B_3 & 0 & \lambda_2\lambda_3 B_1 \\
\lambda_3 E_3 & \lambda_1\lambda_3 B_2 & -\lambda_2\lambda_3 B_1 0 &
\end{pmatrix}
$$

Then the Lorentz force law (7.27) becomes

$$\begin{pmatrix} \left(\frac{Dv}{d\tau}\right)_0 \\ \left(\frac{Dv}{d\tau}\right)_1 \\ \left(\frac{Dv}{d\tau}\right)_2 \\ \left(\frac{Dv}{d\tau}\right)_3 \end{pmatrix} = \frac{q}{mc} \begin{pmatrix} 0 & -\lambda_1 E_1 & -\lambda_2 E_2 & -\lambda_3 E_3 \\ \lambda_1 E_1 & 0 & \lambda_1\lambda_2 B_3 & -\lambda_1\lambda_3 B_2 \\ \lambda_2 E_2 & -\lambda_1\lambda_2 B_3 & 0 & \lambda_2\lambda_3 B_1 \\ \lambda_3 E_3 & \lambda_1\lambda_3 B_2 & -\lambda_2\lambda_3 B_1 & 0 \end{pmatrix} \begin{pmatrix} \dot{y}^0 \\ \dot{y}^1 \\ \dot{y}^2 \\ \dot{y}^3 \end{pmatrix}$$

where $\dot{y}^i = \partial y^i / \partial \tau$. Further we have that

$$\left(\frac{Dv}{d\tau}\right)_k = \sum_{j=0}^{3} g_{kj} \ddot{y}^j + \sum_{i,j=0}^{3} [ij,k]\, \dot{y}^i \dot{y}^j$$

where

$$[ij,k] = \frac{1}{2}\left(\frac{\partial}{\partial y^j}(\lambda_i^2)\delta_{ik} + \frac{\partial}{\partial y^i}(\lambda_j^2)\delta_{jk} - \frac{\partial}{\partial y^k}(\lambda_i^2)\delta_{ij}\right)$$

where we defined $\lambda_0 = 1$. As an example we can take cylindrical coordinates $y = (ct, r, \theta, z) = (y^0, y^1, y^2, y^3)$ relted to the Cartesian coordinates by

$$\begin{aligned} x^1 &= r\cos\theta \\ x^2 &= r\sin\theta \\ x^3 &= z \end{aligned} \tag{7.79}$$

We then have

$$g = -dy^0 \otimes dy^0 + dr \otimes dr + r^2 d\theta \otimes d\theta + dz \otimes dz$$

So $\lambda_0 = \lambda_1 = \lambda_3 = 1$ and $\lambda_2 = r$. Denoting

$$(E_1, E_2, E_3) = (E_r, E_\theta, E_z) \quad , \quad (B_1, B_2, B_3) = (B_r, B_\theta, B_z)$$

and using the only non-vanishing Christoffel symbols

$$[r\theta, \theta] = [\theta r, \theta] = r \quad , \quad [\theta\theta, r] = -r$$

we have

$$\begin{pmatrix} -\ddot{y}^0 \\ \ddot{r} - r\dot{\theta}^2 \\ r^2\ddot{\theta} + 2r\dot{r}\dot{\theta} \\ \ddot{z} \end{pmatrix} = \frac{q}{mc} \begin{pmatrix} 0 & -E_r & -rE_\theta & -E_z \\ E_r & 0 & rB_z & -B_\theta \\ rE_\theta & -rB_z & 0 & rB_r \\ E_z & B_\theta & -rB_r & 0 \end{pmatrix} \begin{pmatrix} \dot{y}^0 \\ \dot{r} \\ \dot{\theta} \\ \dot{z} \end{pmatrix}$$

Using $y^0 = ct$ this yields the equations

$$\begin{aligned} \ddot{t} &= \frac{q}{mc^2}(\dot{r}E_r + r\dot{\theta}E_\theta + \dot{z}E_z) \\ \ddot{r} - r\dot{\theta}^2 &= \frac{q}{m}E_r\dot{t} + \frac{q}{mc}(rB_z\dot{\theta} - B_\theta\dot{z}) \\ r^2\ddot{\theta} + 2r\dot{r}\dot{\theta} &= \frac{q}{m}rE_\theta\dot{t} + \frac{q}{mc}(rB_r\dot{z} - rB_z\dot{r}) \\ \ddot{z} &= \frac{q}{m}E_z\dot{t} + \frac{q}{mc}(B_\theta\dot{r} - rB_r\dot{\theta}) \end{aligned}$$

Chapter 8

The general solution of Maxwell's equations

We show that the electromagnetic field tensor can be written as the exterior derivative of a four-potential. This four potential is not unique but determined up to a gauge transformation. The equation of motion of the four-potential is given by a second order differential equations and we therefore give a discussion of the d'Alembert operator in general coordinates. We discuss the free space solutions of Maxwell's equations that are given by electromagnetic waves and finally give the general solution of the Maxwell equations in terms of given charges and currents.

8.1 Gauge invariance and vector potential

We have seen that the electromagnetic field satisfies the condition $dF = 0$. Given the fact that for any p-form ω we have that $d^2\omega = 0$ it is natural to ask whether there is a one-form A such that

$$F = dA \tag{8.1}$$

and hence $dF = d^2 A = 0$. Differential forms ω with the property that $d\omega = 0$ are called *closed*, and the ones with the property that there is a β such that $\omega = d\beta$ are called *exact*. We know that F is closed, but the question is whether it also is exact. This question is answered by so-called Poincaré's lemma, which states that if a p-form vanishes in a so-called star-shaped region then there is a $(p-1)$-form β such that $d\beta = \omega$. In order not to interrupt the discussion this lemma is proven in Appendix C. Now since dF vanishes everywhere it follows from this lemma that there is a one-form

$$A = A_0\, dx^0 + A_1\, dx^1 + A_2\, dx^2 + A_3\, dx^3 \tag{8.2}$$

with the property (8.1). Insering this expression into $F = dA$ we find that

$$\begin{aligned} F &= \sum_{\mu<\nu}^{3} F_{\mu\nu}\, dx^\mu \wedge dx^\nu = d\left(\sum_{\nu=0}^{3} A_\nu dx^\nu\right) = \sum_{\mu,\nu=0}^{3} \frac{\partial A_\nu}{\partial x^\mu}\, dx^\mu \wedge dx^\nu \\ &= \sum_{\mu<\nu}^{3} \left(\frac{\partial A_\nu}{\partial x^\mu} - \frac{\partial A_\mu}{\partial x^\nu}\right) dx^\mu \wedge dx^\nu \end{aligned}$$

and therefore we find that

$$F_{\mu\nu} = \frac{\partial A_\nu}{\partial x^\mu} - \frac{\partial A_\mu}{\partial x^\nu} \tag{8.3}$$

Since

$$F_{0\nu} = -E_\nu = \frac{\partial A_\nu}{\partial x^0} - \frac{\partial A_0}{\partial x^\nu} \quad (\nu = 1,2,3)$$

it follows that the electric field is given by

$$\mathbf{E} = \nabla A_0 - \frac{1}{c}\frac{\partial \mathbf{A}}{\partial t}$$

where we defined $\mathbf{A} = (A_1, A_2, A_3)$. Further since

$$B_1 = F_{23} = \frac{\partial A_3}{\partial x^2} - \frac{\partial A_2}{\partial x^3} \quad , \quad B_2 = -F_{13} = \frac{\partial A_1}{\partial x^3} - \frac{\partial A_3}{\partial x^1} \quad , \quad B_3 = F_{12} = \frac{\partial A_2}{\partial x^1} - \frac{\partial A_1}{\partial x^2}$$

it follows that

$$\mathbf{B} = \nabla \times \mathbf{A}$$

The quantity A_0 is usually denoted $A_0 = -\phi$ (the minus sign is explained by the fact that raising the indices makes it vanish) such that Eq.(8.2) is written

$$A = -\phi\, dx^0 + A_1\, dx^1 + A_2\, dx^2 + A_3\, dx^3 = (-\phi, \mathbf{A})$$

where A is called the four-potential and $\mathbf{A}$ is called the *vector potential*. In terms of these quantities the electric and magnetic fields are then given by

$$\mathbf{E} = -\nabla\phi - \frac{1}{c}\frac{\partial \mathbf{A}}{\partial t} \tag{8.4}$$

$$\mathbf{B} = \nabla \times \mathbf{A} \tag{8.5}$$

If we write $F = dA$ the equation $dF = 0$ is automatically satisfied. The remaining Maxwell equations are then given by

$$\frac{4\pi}{c}\, J = \star d \star F = \star d \star dA \tag{8.6}$$

This gives an equation for the four-potential A. We can not expect a unique solution to this equation, even when we specify initial conditions. This is because if A is a solution to this equation then also $A' = A + d\Lambda$ is a solution where Λ is an arbitrary function. This follows simply from

$$dA' = d(A + d\Lambda) = dA + d^2\Lambda = dA$$

since $d^2\Lambda = 0$ for any function Λ. It then follows that

$$F' = dA' = dA = F$$

This means that the electric and magnetic fields are invariant under the transformation $A \to A + d\Lambda$. Such a transformation is called a *gauge transformation* and we therefore see that the electromagnetic field tensor is *gauge invariant*. In component notation the transformation $A' = A + d\Lambda$ is written as

$$\phi' = \phi - \frac{1}{c}\frac{\partial \Lambda}{\partial t}$$

$$\mathbf{A}' = \mathbf{A} + \nabla\Lambda$$

It is readily seen that these transformations indeed leave Eqs.(8.4) and (8.5) invariant. We will now investigate Eq.(8.6) a bit further. We see that is involves two d operators and therefore two derivatives. We will now rewrite in a form that involves second covariant derivatives. Let us therefore briefly discuss the topic of taking second derivatives of tensors. We start by pointing out an issue which can lead to confusion when we use an improper notation. Recall that if we have a tensor A of type (p,q) then the covariant derivative ∇ produces a new tensor ∇A of type $(p+1,q)$. Applying the operator twice leads to the tensor $\nabla\nabla A$ of type $(p+2,q)$. For example, if A is a vector of the form

$$A = \sum_i A^i \frac{\partial}{\partial x^j}$$

then

$$\nabla A = \sum_{i,j} A^i_{;j}\, dx^j \otimes \frac{\partial}{\partial x^i} \tag{8.7}$$

where

$$A^i_{;j} = \frac{\partial A^i}{\partial x^j} + \sum_l A^l \Gamma^i_{jl}$$

are the components of a tensor of type $(1,1)$. Taking the covariant derivative again we obtain

$$\nabla\nabla A = \sum_{i,j,k} (A^i_{;j})_{;k}\, dx^k \otimes dx^j \otimes \frac{\partial}{\partial x^i} = \sum_{i,j,k} (\nabla\nabla A)^i_{kj}\, dx^k \otimes dx^j \otimes \frac{\partial}{\partial x^i} \tag{8.8}$$

where

$$(\nabla\nabla A)^i_{kj} = (A^i_{;j})_{;k} = \frac{\partial A^i_{;j}}{\partial x^k} - \sum_l A^i_{;l}\Gamma^l_{kj} + \sum_l A^l_{;j}\Gamma^i_{kl} \tag{8.9}$$

where we simply used Eq.(5.49) for the differentiation of a tensor of type $(1,1)$. Let us now do a similar but relating thing. We take again the vector A but consider is derivative in the direction of the vector v. This gives a new vector B given by

$$B = \nabla_v A = \sum_{i,j} A^i_{;j} v^j \frac{\partial}{\partial x^i} = \sum_i B_i \frac{\partial}{\partial x^i}$$

where

$$B^i = \sum_j A^i_{;j} v^j$$

We can continue to differentiate B in the direction of another vector field w. This gives a new vector C given by

$$C = \nabla_w B = \sum_{i,k} B^i_{;k} w^k \frac{\partial}{\partial x^i} = \sum_i C_i \frac{\partial}{\partial x^i}$$

where

$$(\nabla_w \nabla_v A)^i = C^i = \sum_k B^i_{;k} w^k = \sum_{j,k} (A^i_{;j} v^j)_{;k} w^k \tag{8.10}$$

where

$$B^i_{;k} = \frac{\partial B^i}{\partial x^k} + \sum_l B^l \Gamma^i_{kl} = \frac{\partial}{\partial x^k}(\sum_j A^i_{;j} v^j) + \sum_{j,l} A^l_{;j} v^j \Gamma^i_{kl} \tag{8.11}$$

If we now take $v = \partial/\partial x^j$ and $w = \partial/\partial x^k$ we find from Eq.(8.10) and (8.11) that

$$(\nabla_k \nabla_j A)^i = \frac{\partial A^i_{;j}}{\partial x^k} + \sum_l A^l_{;j}\Gamma^i_{kl} \tag{8.12}$$

The confusion thing is that the right hand side is not equal to the expression (8.9) as the last term in Eq.(8.10) may suggest, since here we take the covariant derivative ∇_k of the vector B rather than a tensor of type $(1,1)$. The index j in the expression

$$(\nabla_k \nabla_j A)^i = B^i_{;k} = (A^i_{;j})_{;k} \tag{8.13}$$

is not regarded as tensor index when taking the covariant derivative ∇_k and therefore the right hand side of this expression is not the same as (8.9) . We see from Eq.(8.9) and Eq.(8.12) that

$$(A^i_{;j})_{;k} = (\nabla_k \nabla_j A)^i - \sum_l A^i_{;l} \Gamma^l_{kj}$$

where the left hand side now has the meaning of a tensor (and j is now regarded as a tensor index.. We see from this example that it os not a good idea to use the notation on the right hand side of Eq.(8.13) for the second covariant derivative as it may lead to confusion in calculations if one does not keep track on the meaning of the indices. Therefore, we will introduce the notation that for a given tensor of type (p,q) we will always regard

$$\left(A^{j_1 \ldots j_q}_{i_1 \ldots i_p; l} \right)_{;k} = A^{j_1 \ldots j_q}_{i_1 \ldots i_p; l; k}$$

as components of the tensor $\nabla \nabla A$ of type $(p+2, q)$, whereas the components of the tensor $\nabla_k \nabla_l A$ of type (p,q) will be denoted by

$$(\nabla_k \nabla_l A)^{j_1 \ldots j_q}_{i_1 \ldots i_p} = \nabla_k \nabla_l A^{j_1 \ldots j_q}_{i_1 \ldots i_p}$$

where we remove the brackets to make the notation less busy. We should however keep in mind that we are not differentiating the components themselves, but always take the components after differentiation. In the following we will often raise indices of variables with respect we differentiate and we therefore define

$$\nabla^l A^{j_1 \ldots j_q}_{i_1 \ldots i_p} = \sum_m g^{lm} \, \nabla_m A^{j_1 \ldots j_q}_{i_1 \ldots i_p}$$

or more compactly

$$\nabla^l = \sum_m g^{lm} \, \nabla_m$$

Using this definition we can derive useful relations such as

$$\sum_k \nabla^k A_k = \sum_{i,j,k} g^{ki} \nabla_i (g_{kj} A^j) = \sum_{i,j,k} g^{ki} g_{kj} \nabla_i A^j = \sum_{i,j} \delta^i_j \nabla_i A^j = \sum_j \nabla_j A^j$$

where we used that $\nabla_i g_{kj} = 0$. After this intermezzo we can now go back to the electromagnetic tensor. We write this as

$$\begin{aligned} F_{\mu\nu} &= \frac{\partial A_\nu}{\partial x^\mu} - \frac{\partial A_\mu}{\partial x^\nu} = \left(\frac{\partial A_\nu}{\partial x^\mu} - \sum_\rho A_\rho \Gamma^\rho_{\mu\nu} \right) - \left(\frac{\partial A_\mu}{\partial x^\nu} - \sum_\rho A_\rho \Gamma^\rho_{\nu\mu} \right) \\ &= A_{\nu;\mu} - A_{\mu;\nu} = \nabla_\mu A_\nu - \nabla_\nu A_\mu \end{aligned}$$

where we used the symmetry of the Christoffel symbols. By raising the indices we then find that

$$F^{\mu\nu} = \nabla^\mu A^\nu - \nabla^\nu A^\mu \tag{8.14}$$

Then we can write Maxwell's equations as

$$\frac{4\pi}{c}\,j^\mu = \sum_\nu \nabla_\nu F^{\mu\nu} = \sum_\nu \nabla_\nu (\nabla^\mu A^\nu - \nabla^\nu A^\mu)$$

By lowering the index μ we obtain

$$\sum_\nu \nabla_\nu \nabla_\mu A^\nu - \nabla_\nu \nabla^\nu A_\mu = \frac{4\pi}{c}\,J_\mu \tag{8.15}$$

This is the more explicit version of Eq.(8.6). As we noted this equation does not have a unique solution due to the gauge freedom that we have and we therefore need to fix a gauge. More precisely we need to make a choice for the scalar function Λ. We make the so-called *Lorenz gauge* and require that

$$\sum_{\nu=0}^{3} \nabla_\nu A^\nu = \sum_{\nu=0}^{3} A^\nu_{;\nu} = 0 \tag{8.16}$$

This is always possible. Suppose namely that we had a four-potential A' with the property

$$\sum_{\nu=0}^{3} \nabla_\nu A'^\nu = f$$

where f is some function. Then by choosing Λ such that it satisfies the differential equation

$$\sum_{\nu=0}^{3} \nabla_\nu (d\Lambda)^\nu = \sum_{\mu,\nu=0}^{3} \frac{1}{\sqrt{|g|}} \left(\sqrt{|g|} g^{\nu\mu} \frac{\partial \Lambda}{\partial x^\mu} \right) = f \tag{8.17}$$

we find that $A = A' - d\Lambda$ exactly satisfies the gauge of Eq.(8.16). This still does not fix all gauge freedom as this equation as to any solution Λ of this equation we can add a homogeneous solution Λ_0 satisfying $\sum_\nu \nabla_\nu (d\Lambda_0)^\nu = 0$ such that $\Lambda_0 + \Lambda$ also satisfies the equation above. If we use the Lorenz gauge we can write Eq.(8.15) as

$$\begin{aligned} \frac{4\pi}{c}\,J_\mu &= \sum_{\nu=0}^{3} (-\nabla_\nu \nabla^\nu A_\mu + (\nabla_\nu \nabla_\mu - \nabla_\mu \nabla_\nu) A^\nu + \nabla_\mu \nabla_\nu A^\nu) \\ &= \sum_{\nu=0}^{3} (-\nabla_\nu \nabla^\nu A_\mu + [\nabla_\nu, \nabla_\mu] A^\nu) \end{aligned} \tag{8.18}$$

where we defined the commutator

$$[\nabla_\nu, \nabla_\mu] = \nabla_\nu \nabla_\mu - \nabla_\mu \nabla_\nu$$

In general this commutator does not vanish since covariant derivatives do not commute due to the presence of the Christoffel symbols. However, in the special case of a standard Minkowskian metric we have simply

$$\nabla_l A^{j_1 \dots j_q}_{i_1 \dots i_p} = \frac{\partial A^{j_1 \dots j_q}_{i_1 \dots i_p}}{\partial x^l} \tag{8.19}$$

in which case we can identify ∇_l with the simple differential operator $\partial_l = \partial / \partial x^l$, for which

$$[\partial_\nu, \partial_\mu] = 0$$

In that case Eq.(8.18) attains the form

$$\begin{aligned}\frac{4\pi}{c}\,J_\mu &= \sum_{\nu=0}^{3} -\partial_\nu \partial^\nu A_\mu = \sum_{\nu,\rho=0}^{3} g^{\nu\rho} \frac{\partial^2 A_\mu}{\partial x^\nu \partial x^\rho} \\ &= -\left[-\frac{\partial^2}{\partial x^{0^2}} + \frac{\partial^2}{\partial x^{1^2}} + \frac{\partial^2}{\partial x^{2^2}} + \frac{\partial^2}{\partial x^{3^2}} \right] A_\mu = -\Box\, A_\mu \end{aligned} \tag{8.20}$$

where we defined the d'Alembertian $\Box$ by

$$\Box = -\frac{1}{c^2}\frac{\partial^2}{\partial t^2} + \frac{\partial^2}{\partial x^{1^2}} + \frac{\partial^2}{\partial x^{2^2}} + \frac{\partial^2}{\partial x^{3^2}}$$

We will in detail investigate the solution to Eq.(8.20) but before we do this we will first investigate what happens if we do not assume that the commutator in Eq.(8.18) vanishes. Let us simply calculate how the commutator looks in given coordinates when we act on a vector. We have

$$\begin{aligned}[\nabla_k, \nabla_j] A^i &= \nabla_k \nabla_j A^i - \nabla_j \nabla_k A^i \\ &= \left(\frac{\partial A^i_{;j}}{\partial x^k} + \sum_l^n A^l_{;j} \Gamma^i_{kl} \right) - \left(\frac{\partial A^i_{;k}}{\partial x^j} + \sum_l^n A^l_{;k} \Gamma^i_{jl} \right) \\ &= \frac{\partial}{\partial x^k} \left(\frac{\partial A^i}{\partial x^j} + \sum_m^n A^m \Gamma^i_{jm} \right) + \sum_l^n \left(\frac{\partial A^l}{\partial x^j} + \sum_m^n A^m \Gamma^l_{jm} \right) \Gamma^i_{kl} \\ &\quad - \frac{\partial}{\partial x^j} \left(\frac{\partial A^i}{\partial x^k} + \sum_m^n A^m \Gamma^i_{km} \right) - \sum_l^n \left(\frac{\partial A^l}{\partial x^k} + \sum_m^n A^m \Gamma^l_{km} \right) \Gamma^i_{jl} \\ &= -\sum_m^n A^m R^i_{mjk} \end{aligned} \tag{8.21}$$

where we defined (using the symmetry $\Gamma^i_{jk} = \Gamma^i_{kj}$ of the Christoffel symbols)

$$R^i_{mjk} = \frac{\partial \Gamma^i_{mk}}{\partial x^j} - \frac{\partial \Gamma^i_{mj}}{\partial x^k} + \sum_l^n (\Gamma^l_{mk}\Gamma^i_{lj} - \Gamma^l_{mj}\Gamma^i_{lk})$$

which we recognize as the Riemann curvature tensor of Eq.(4.56). Eq.(8.21) is a special case of the following general relation that can be proven by a similar straightforward calculation

$$[\nabla_k, \nabla_l] A^{j_1\ldots j_q}_{i_1\ldots i_p} = \sum_{r=1}^{p}\sum_{\nu=1}^{n} A^{j_1\ldots j_q}_{i_1\ldots i_{r-1}\nu i_{r+1}\ldots i_p} R^\nu_{i_r kl} - \sum_{s=1}^{q}\sum_{\nu=1}^{n} A^{j_1\ldots j_{s-1}\nu j_{s+1}\ldots j_q}_{i_1\ldots i_p} R^{j_s}_{\nu kl} \tag{8.22}$$

Using Eq.(8.21) we find that Eq.(8.18) becomes

$$\begin{aligned}\frac{4\pi}{c}\,J_\mu &= \sum_{\nu=0}^{3} (-\nabla_\nu \nabla^\nu A_\mu + [\nabla_\nu, \nabla_\mu] A^\nu) \\ &= \sum_{\nu=0}^{3} (-\nabla_\nu \nabla^\nu A_\mu - \sum_\rho A^\rho R^\nu_{\rho\mu\nu}) = -\sum_{\nu=0}^{3} \nabla_\nu \nabla^\nu A_\mu + \sum_\rho A^\rho R_{\rho\mu} \end{aligned} \tag{8.23}$$

where we defined

$$R_{\rho\mu} = -\sum_\nu R^\nu_{\rho\mu\nu}$$

which from Eq.(4.63) we recognize as the Ricci tensor. As discussed in Chapter 4 the significance of the Riemann curvature tensor is that it vanishes only when there exist a coordinate transformation which makes the metric diagonal. In our case this implies that there is a coordinate transformation which makes the metric of the form

$$g = -dx^0 \otimes dx^0 + dx^1 \otimes dx^1 + dx^2 \otimes dx^2 + dx^3 \otimes dx^3$$

This still includes a wide variety of physical situations such as the rotating disc of Chapter 2 or any other non-inertial frame in flat space-time. However, in general gravity fields the use of Eq.(8.23) becomes necessary although the vanishing of the Ricci tensor is a weaker constraint than the vanishing of the Riemann tensor. For instance, in the Schwarzschild solution for the black hole the Ricci tensor vanishes. We conclude that if we can neglect the curvature of space due to mass-energy distributions then we can put $R_{\rho\mu} = 0$ and then Eq.(8.23) becomes

$$\frac{4\pi}{c} J_\mu = -\sum_{\nu=0}^{3} \nabla_\nu \nabla^\nu A_\mu \tag{8.24}$$

This equation is still valid in general coordinate systems but only within the class of metrics $g_{\mu\nu}$ that can be obtained from the Minkowski metric by a coordinate transformation. If we are interested in electromagnetic fields in curved space-time we can always add the last term in Eq.(8.23).

8.2 The Laplacian and d'Alembertian

What we will show now is that we can find an elegant expression for both the d'Alembertian and the gauge-fixing condition using differential forms. In Eq.(7.50) we saw that the Laplacian operator Δ on a function f is given by

$$\Delta f = \nabla^2 f = \star d \star df$$

If we, however, want to generalize this definition to vectors rather than scalar functions the expression in terms of differential forms changes. For instance, in three dimensions we have the well-known identity

$$-\Delta \mathbf{A} = -\nabla^2 \mathbf{A} = \nabla \times (\nabla \times \mathbf{A}) - \nabla(\nabla \cdot \mathbf{A}) \tag{8.25}$$

as can be readily seen by a short calculation

$$\begin{aligned} \nabla \times (\nabla \times \mathbf{A})_i - \nabla(\nabla \cdot \mathbf{A})_i &= \sum_{j,k} \epsilon_{ijk} \partial_j (\nabla \times A)_k - \partial_i \sum_j \partial_j A_j \\ = \sum_{j,k,l,m} \epsilon_{ijk}\epsilon_{klm} \partial_j \partial_l A_m - \sum_j \partial_i \partial_j A_j &= \sum_{j,l,m} (\delta_{il}\delta_{jm} - \delta_{im}\delta_{jl}) \partial_j \partial_l A_m - \sum_j \partial_i \partial_j A_j \\ = \sum_j \partial_j \partial_i A_j - \partial_j \partial_j A_i - \sum_j \partial_i \partial_j A_j &= -\sum_j \partial_j^2 A_i = -\Delta A_i \end{aligned}$$

If A is a one-form then we saw that curl is given by the operator $\star d$ and therefore

$$\nabla \times (\nabla \times \mathbf{A}) = \star d \star dA$$

we further noted that d represents the gradient when acting on a scalar function and $\star d \star$ represents the divergence when acting on a one-form. We therefore find

$$\nabla(\nabla \cdot \mathbf{A}) = d \star d \star A$$

Therefore from Eq.(8.25) we find that form a one-form A in three dimensions with standard metric we have

$$-\Delta A = (\star\, d \star d - d \star d \star)\, A \tag{8.26}$$

This representation of the Laplacian Δ also seems to work for scalar functions apart from a sign since

$$d \star f = d(\Omega f) = 0$$

since the differential of the volume form Ω vanishes. It therefore appears that Eq.(8.26) could be generalization of the Laplacian for general p-forms, but we need to fix the signs appropriately. Moreover it turns out that the relative minus sign between the two terms is dependent on the number of dimensions. In Minkowski space it must become a plus sign in order to reproduce the d'Alembertian. To find the right sign we must apply the operators $\star\, d \star d$ and $d \star d \star$ on a general p-form and compare the results. Let us start by considering the operator $\star\, d \star$ acting on a general p-form ω. Let us go back to Eq.(7.61) and write it in the form

$$\begin{aligned}(\star\, d \star \omega)_{m_1 \ldots m_{p-1}} &= \frac{\mathrm{sign}(g)}{\sqrt{|g|}} (-1)^{np+n} \sum_{k, r_1 \ldots r_{p-1}}^{n} \frac{\partial}{\partial x^k} (\sqrt{|g|}\, \omega^k{}_{m_1 \ldots m_{p-1}}) \\ &= \mathrm{sign}(g) (-1)^{np+n} \sum_{k}^{n} \omega^k{}_{m_1 \ldots m_{p-1};k}\end{aligned} \tag{8.27}$$

where we again used that lowering of indices commutes with covariant differentiation. We can also turn the equation around and write

$$-\sum_{k}^{n} \omega^k{}_{m_1 \ldots m_{p-1};k} = (d^\dagger \omega)_{m_1 \ldots m_{p-1}} \tag{8.28}$$

where we define the so-called *co-differential operator* $d^\dagger$ as

$$d^\dagger = \mathrm{sign}(g) (-1)^{np+n+1} \star d \star \tag{8.29}$$

It turns out that $d^\dagger$ is the adjoint of d in a suitably chosen inner product but we will not need this feature here. We will define it more precisely later when we will look at the action principle for the Maxwell equations. Just like the d operator the operator $d^\dagger$ has the property $(d^\dagger)^2 = 0$, as follows immediately from

$$(\star d \star)^2 = \star d \star \star d \star = \pm \star d^2 \star = 0$$

The generalization of the Laplacian Δ or the d'Alembertian (which is usually written as $\Box$ instead) on general p-forms is then given by

$$-\Delta = d\, d^\dagger + d^\dagger d$$

Before we show this we will first check that we get back Eq.(8.26) for one-forms in three dimensions with an Euclidean metric. If A is a one-form then dA is a two-form and we find

$$\begin{aligned}(d\, d^\dagger + d^\dagger d) A &= (-1)^{3\cdot 1+3+1} d \star d \star A + (-1)^{3\cdot 2+3+1} \star d \star dA \\ &= (\star\, d \star d - d \star d \star)\, A\end{aligned}$$

which agrees with Eq.(8.26). You can check yourself that we also get the right result for the Laplacian when we act on a function. Let us now act with $d\, d^\dagger$ and $d^\dagger d$ on general p-forms. We start with $d^\dagger d$. For a p-form

$$\omega = \frac{1}{p!} \sum_{i_1 \ldots i_p}^{n} \omega_{i_1 \ldots i_p} dx^{i_1} \wedge \ldots \wedge dx^{i_p}$$

we have

$$d\omega = \frac{1}{p!}\sum_{i_1\ldots i_p}^{n} \frac{\partial \omega_{i_1\ldots i_p}}{\partial x^k} dx^k \wedge dx^{i_1} \wedge \ldots \wedge dx^{i_p}$$

Due to the symmetry of the Christoffel symbols we can replace $\partial_k \omega_{i_1\ldots i_p}$ by $\omega_{i_1\ldots i_p;k}$. We can therefore write

$$\begin{aligned} d\omega &= \frac{1}{p!}\sum_{i_1\ldots i_p}^{n} \nabla_k\, \omega_{i_1\ldots i_p}\, dx^k \wedge dx^{i_1} \wedge \ldots \wedge dx^{i_p} \\ &= \frac{1}{p!(p+1)!}\sum_{\substack{i_1\ldots i_p,k \\ j_1\ldots j_{p+1}}}^{n} \delta^{k\, i_1\ldots i_p}_{j_1\ldots j_{p+1}} \nabla_k\, \omega_{i_1\ldots i_p}\, dx^{j_1} \wedge \ldots \wedge dx^{j_{p+1}} \\ &= \frac{1}{(p+1)!}\sum_{j_1\ldots j_{p+1}}^{n} (d\omega)_{j_1\ldots j_{p+1}}\, dx^{j_1} \wedge \ldots \wedge dx^{j_{p+1}} \end{aligned}$$

where we defined

$$(d\omega)_{j_1\ldots j_{p+1}} = \frac{1}{p!}\sum_{k,i_1\ldots i_p}^{n} \delta^{k\, i_1\ldots i_p}_{j_1\ldots j_{p+1}} \nabla_k\, \omega_{i_1\ldots i_p} \tag{8.30}$$

From Eq.(8.28) we then have

$$\begin{aligned} (d^\dagger d\omega)_{r_1\ldots r_p} &= -\sum_{k}^{n} \nabla^k (d\omega)_{k\, r_1\ldots r_p} \\ &= -\frac{1}{p!}\sum_{k,l,i_1\ldots i_p}^{n} \delta^{l\, i_1\ldots i_p}_{k\, r_1\ldots r_p} \nabla^k \nabla_l\, \omega_{i_1\ldots i_p} \\ &= -\frac{1}{p!}\sum_{k,i_1\ldots i_p}^{n} \delta^{k\, i_1\ldots i_p}_{k\, r_1\ldots r_p} \nabla^k \nabla_k\, \omega_{i_1\ldots i_p} - \frac{1}{p!}\sum_{k\neq l}^{n}\sum_{i_1\ldots i_p}^{n} \delta^{l\, i_1\ldots i_p}_{k\, r_1\ldots r_p} \nabla^k \nabla_l\, \omega_{i_1\ldots i_p} \\ &= -\sum_{k}^{n} \nabla^k \nabla_k\, \omega_{r_1\ldots r_p} + \frac{1}{(p-1)!}\sum_{l}^{n}\sum_{i_1,\ldots,i_p}^{n} \delta^{l\, i_2\ldots i_p}_{r_1\ldots r_p} \nabla^{i_1} \nabla_l\, \omega_{i_1\ldots i_p} \end{aligned} \tag{8.31}$$

where in the evaluation of the last term we used that the index k must be equal to one of the indices $(i_1 \ldots i_p)$ but each of the p possible choices $k = i_m$ leads to the same sum since both δ and ω are anti-symmetric. We therefore simply made the choice $k = i_1$ and multiplied with p. We further used that

$$\delta^{l\, i_1\, i_2\ldots i_p}_{i_1\, r_1\ldots r_p} = -\delta^{i_1\, l\, i_2\ldots i_p}_{i_1\, r_1\ldots r_p} = -\delta^{l\, i_2\ldots i_p}_{r_1\ldots r_p}$$

and the fact that $l \neq i_1$. We now proceed to calculate $d\, d^\dagger$. Since $d^\dagger\omega$ is a $(p-1)$-form we have from Eq.(8.30) that

$$\begin{aligned} (d\, d^\dagger\omega)_{r_1\ldots r_p} &= \frac{1}{(p-1)!}\sum_{k\, i_1\ldots i_{p-1}}^{n} \delta^{k\, i_1\ldots i_{p-1}}_{r_1\ldots r_p} \nabla_k (d^\dagger\omega)_{i_1\ldots i_{p-1}} \\ &= -\frac{1}{(p-1)!}\sum_{l,k,i_1\ldots i_{p-1}}^{n} \delta^{k\, i_1\ldots i_{p-1}}_{r_1\ldots r_p} \nabla_k \nabla^l\, \omega_{l\, i_1\ldots i_{p-1}} \end{aligned}$$

If we add this equation to Eq.(8.31) we then obtain that

$$([d^\dagger d + d\,d^\dagger]\omega)_{r_1\ldots r_p} = -\sum_k^n \nabla^k \nabla_k\, \omega_{r_1\ldots r_p} + \frac{1}{(p-1)!}\sum_{k,l}^n \sum_{i_2,\ldots,i_p}^n \delta^{l\,i_2\ldots i_p}_{r_1\ldots r_p}[\nabla^k, \nabla_l]\,\omega_{k\,i_2\ldots i_p}$$

$$= -\sum_k^n \nabla^k \nabla_k\, \omega_{r_1\ldots r_p} + \frac{1}{(p-1)!}\sum_{k,l}^n \sum_{i_2,\ldots,i_p}^n \delta^{l\,i_2\ldots i_p}_{r_1\ldots r_p}[\nabla_k, \nabla_l]\,\omega^k{}_{i_2\ldots i_p} \tag{8.32}$$

This is our final result. If ω is equal to the one-form A this equation gives

$$[d^\dagger d + d\,d^\dagger]A_\mu = -\sum_\nu \nabla^\nu \nabla_\nu A_\mu + \sum_\nu [\nabla_\nu, \nabla_\mu]A^\nu$$

On the right hand side we have obtained exactly the same as in Eq.(8.18). We can therefore equivalently write

$$(d^\dagger d + d\,d^\dagger)A = \frac{4\pi}{c}J \tag{8.33}$$

Using Eq.(8.22) we could have written out Eq.(8.32) further in terms of the Riemann curvature tensor. If we are working in the absence of gravitational fields we can forget about these terms since in that case they are identically zero. In that case Eq.(8.32) simply gives

$$([d^\dagger d + d\,d^\dagger]\omega)_{r_1\ldots r_p} = -\sum_k^n \nabla^k \nabla_k\, \omega_{r_1\ldots r_p}$$

We then indeed see that $d^\dagger d + d\,d^\dagger$ gives the required generalization of the Laplacian. If we now go back to four space-time dimensions and use standard Minkowskian coordinates then we see that

$$\sum_k^n \nabla^k \nabla_k\, \omega_{r_1\ldots r_p} = \left[-\frac{\partial^2}{\partial x^{0^2}} + \frac{\partial^2}{\partial x^{1^2}} + \frac{\partial^2}{\partial x^{2^2}} + \frac{\partial^2}{\partial x^{3^2}}\right]\omega_{r_1\ldots r_p}$$

We therefore define the generalization of the d'Alembertian as

$$\Box = -(d^\dagger d + d\,d^\dagger) \tag{8.34}$$

In four-dimensional space-time we see from the definition in Eq.(8.29) that

$$d^\dagger = \mathrm{sign}(g)(-1)^{4p+4+1} \star d \star = \star\, d \star$$

for any p-form and therefore

$$-\Box = \star\, d \star d + d \star d \star \tag{8.35}$$

Let us now go back to our starting equation (8.6). We can now write this as

$$\frac{4\pi}{c}\,J = \star\, d \star dA = -\Box A - d \star d \star A \tag{8.36}$$

We can now use the gauge condition

$$\star\, d \star A = d^\dagger A = 0$$

This gauge condition gives

$$d^\dagger A = \star\, d \star A = -\sum_\nu A^\nu_{;\nu} = 0$$

which is precisely the Lorenz gauge condition. Then Eq. (8.36) yields

$$\Box A = -\frac{4\pi}{c}J \tag{8.37}$$

With $\Box$ defined as in Eqs.(8.34) or (8.35) this equation is valid in general space-times. What remains is to study the solutions of this equation.

8.3 The four-potential and Lorentz transformations

Before we study Eq.(8.37) in its generality it will be useful to give some examples of the use of four-potentials. Let us first write the four-potential A in a four-vector form by raising the indices. If we assume standard Minkowskian coordinates then this gives

$$A^\sharp = \phi\frac{\partial}{\partial x^0} + A_1\frac{\partial}{\partial x^1} + A_2\frac{\partial}{\partial x^2} + A_3\frac{\partial}{\partial x^3} \tag{8.38}$$

By similarly raising the indices in Eq.(8.37) we see that the components of this vector satisfy

$$\Box A^\mu = -\frac{4\pi}{c}\, j^\mu$$

where $j = (c\rho, \mathbf{j})$ is the current four-vector. If we write this out in terms of the components of $A^\sharp = (\phi, \mathbf{A})$ we have

$$\Box\phi = -4\pi\rho \tag{8.39}$$

$$\Box\mathbf{A} = -4\pi\mathbf{j} \tag{8.40}$$

Since the d'Alembertian operator $-\Box = \sum_\nu \nabla_\nu\nabla^\nu$ is a Lorentz invariant these equations are valid in any Lorentz frame. We can check this explicitly. Since $A^\sharp$ is a four-vector we have, for instance, under a Lorentz boost in the x^1-direction that

$$\phi' = \frac{\phi - \frac{v}{c}A^1}{\sqrt{1-\frac{v^2}{c^2}}} \tag{8.41}$$

$$A'^1 = \frac{A^1 - \frac{v}{c}\phi}{\sqrt{1-\frac{v^2}{c^2}}} \tag{8.42}$$

$$A'^2 = A^2 \qquad A'^3 = A^3$$

Then with $\gamma = (1 - v^2/c^2)^{-1/2}$ we have

$$\Box\phi' = \gamma(\Box\phi - \frac{v}{c}\Box A^1) = \gamma(-4\pi\rho + \frac{v}{c}\frac{4\pi}{c}j^1) = -\frac{4\pi}{c}(c\rho - \frac{v}{c}j^1) = -\frac{4\pi}{c}c\rho' = -4\pi\rho'$$

$$\Box A'^1 = \gamma(\Box A^1 - \frac{v}{c}\Box\phi) = \gamma(-\frac{4\pi}{c}j^1 + \frac{v}{c}4\pi\rho) = -\frac{4\pi}{c}\gamma(j^1 - v\rho) = -\frac{4\pi}{c}j'^1$$

where we used the transformation laws for the charge and the current of Eqs.(7.7) and (7.8). Let us now give a useful application of these transformations. We consider a static charge distribution $\rho'(x')$ at rest with respect to a Lorentz frame O'. Let this frame move with velocity v along the x^1-axis of another system O. We want to find the electric and magnetic fields with respect to O. Since in O' we have that $\mathbf{j}' = 0$ we can take $\mathbf{A}' = 0$. The equation for ϕ' is then given by (8.39) as

$$-4\pi\rho' = \left[-\frac{1}{c^2}\frac{\partial^2}{\partial t'^2} + \frac{\partial^2}{\partial x'^2} + \frac{\partial^2}{\partial y'^2} + \frac{\partial^2}{\partial z'^2}\right]\phi'$$

where we denote the coordinates in O' by (ct', x', y', z') and similarly in O by (ct, x, y, z). Since the charge distribution is static in O' we can take the potential ϕ' to be time-independent and our equation reduces to

$$\nabla'^2\phi' = -4\pi\rho'$$

which is simply Poisson's equation from electrostatics. The solution to this equation is well-known to be given by

$$\phi'(\mathbf{r}') = \int d\mathbf{r}''\,\frac{\rho'(\mathbf{r}'')}{|\mathbf{r}' - \mathbf{r}''|}$$

where $|\mathbf{r}| = (x^2+y^2+z^2)^{1/2}$ is the three-dimensional length of $\mathbf{r} = (x,y,z)$ and were we used the boundary condition that $\phi' \to 0$ for $|\mathbf{r}'| \to \infty$. To check that this expression indeed satisfies Poisson's equation we simply need the identity

$$\nabla^2 \frac{1}{|\mathbf{r}|} = -4\pi\delta(\mathbf{r}) \tag{8.43}$$

where $\delta(\mathbf{r})$ is the usual three-dimensional delta distribution. Let us now take the simple case that

$$\rho'(\mathbf{r}') = q\,\delta(\mathbf{r}')$$

is a point charge in the origin of system O'. In that case we have that

$$\phi'(\mathbf{r}') = \frac{q}{|\mathbf{r}'|} \tag{8.44}$$

is simply the Coulomb potential of point charge q. By the reciprocal of the Lorentz transformations (i.e. interchanging the primed and unprimed and replacing v by $-v$) of Eq.(8.41) and (8.42) we then find

$$\begin{aligned}
\phi &= \frac{\phi'(\mathbf{r}')}{\sqrt{1-\frac{v^2}{c^2}}} = \frac{q}{\sqrt{1-\frac{v^2}{c^2}}} \frac{1}{\sqrt{x'^2+y'^2+z'^2}} \\
&= \frac{q}{\sqrt{1-\frac{v^2}{c^2}}} \left(\left(\frac{x-vt}{\sqrt{1-\frac{v^2}{c^2}}} \right)^2 + y^2 + z^2 \right)^{-\frac{1}{2}} && (8.45) \\
A_x &= \frac{v}{c} \frac{\phi'(\mathbf{r}')}{\sqrt{1-\frac{v^2}{c^2}}} = \frac{v}{c}\phi && (8.46) \\
A_y &= A_z = 0
\end{aligned}$$

Now we can calculate the electric and magnetic fields from Eqs.(8.4) and (8.5). This gives using Eq.(8.4)

$$\mathbf{E} = \frac{q\gamma}{(\gamma^2(x-vt)^2+y^2+z^2)^{3/2}} \begin{pmatrix} x-vt \\ y \\ z \end{pmatrix} \tag{8.47}$$

whereas Eq.(8.5) gives

$$\begin{aligned}
\mathbf{B} &= \begin{pmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \\ \frac{\partial}{\partial z} \end{pmatrix} \times \begin{pmatrix} \frac{v}{c}\phi \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ \frac{v}{c}\frac{\partial\phi}{\partial z} \\ -\frac{v}{c}\frac{\partial\phi}{\partial y} \end{pmatrix} = \frac{v}{c} \begin{pmatrix} 0 \\ -E_z \\ E_y \end{pmatrix} \\
&= \frac{v}{c} \begin{pmatrix} v \\ 0 \\ 0 \end{pmatrix} \times \begin{pmatrix} E_x \\ E_y \\ E_z \end{pmatrix} = \frac{1}{c}\mathbf{v}\times\mathbf{E} && (8.48)
\end{aligned}$$

We therefore have the following situation

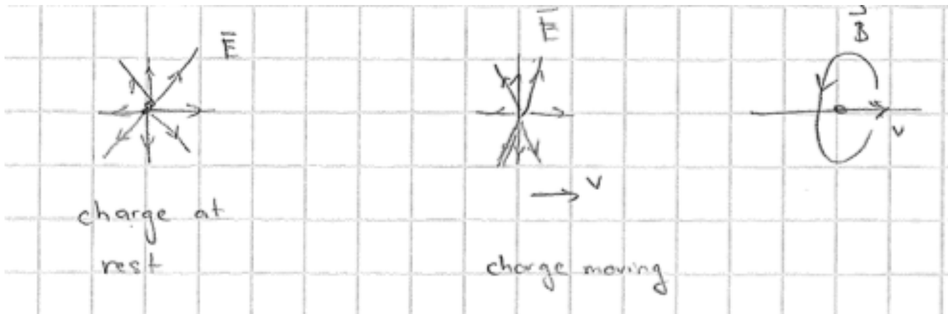

Figure 8.1: Electric and magnetic fields from a uniformly moving charge.

The electric field at $x = vt$ is given by

$$\mathbf{E} = \frac{q\gamma}{(y^2 + z^2)^{3/2}} \begin{pmatrix} 0 \\ y \\ z \end{pmatrix}$$

and is therefore increased by a factor of γ compared to the field in the rest frame. The electric field component E_x along the x-axis ($y = z = 0$) is given by

$$E_x = q\left(1 - \frac{v^2}{c^2}\right)\frac{1}{(x - vt)^2} \tag{8.49}$$

is reduced by a factor γ^2. We will see that for non-uniform motion the particle will start to radiate.

8.4 Solution of Maxwell's equation in free space: electromagnetic waves

Let us now go back to Eq.(8.37) and study it for the case of free space, or the absence of charge or current distributions, $j^\mu = 0$. We will consider a standard Lorentz frame. In that case we have

$$\Box A^\mu = 0$$

If we write this out in terms of components we have

$$\Box\phi = 0 \quad , \quad \Box\mathbf{A} = 0$$

together with the gauge condition

$$\sum_\nu \nabla_\nu A^\nu = \frac{1}{c}\frac{\partial\phi}{\partial t} + \nabla\cdot\mathbf{A} = 0$$

We make use of the fact that this gauge condition still leaves some gauge freedom. If Λ is a function such that

$$-\sum_\nu \nabla_\nu\nabla^\nu\Lambda = \Box\Lambda = 0$$

then the potentials

$$\phi' = \phi - \frac{1}{c}\frac{\partial\Lambda}{\partial t}$$
$$\mathbf{A}' = \mathbf{A} + \nabla\Lambda$$

also satisfy the Lorenz gauge condition since

$$\frac{1}{c}\frac{\partial \phi'}{\partial t} + \nabla \cdot \mathbf{A}' = -\frac{1}{c^2}\frac{\partial^2 \Lambda}{\partial t^2} + \nabla^2 \Lambda + \frac{1}{c}\frac{\partial \phi}{\partial t} + \nabla \cdot \mathbf{A} = \frac{1}{c}\frac{\partial \phi}{\partial t} + \nabla \cdot \mathbf{A} = 0$$

In particular, if ϕ satisfies $\Box \phi = 0$ then we can choose Λ such that

$$\phi = \frac{1}{c}\frac{\partial \Lambda}{\partial t}$$

such that $\phi' = 0$ and the Lorenz gauge condition simply attains the form $\nabla \cdot \mathbf{A}' = 0$. The electric field is then simply given by

$$\mathbf{E} = -\nabla \phi' - \frac{1}{c}\frac{\partial \mathbf{A}'}{\partial t} = -\frac{1}{c}\frac{\partial \mathbf{A}'}{\partial t}$$

If we drop the primes again, we conclude that the equations to be solved are therefore given by

$$\Box \mathbf{A} = 0 \quad , \quad \nabla \cdot \mathbf{A} = 0$$

from which the electric and magnetic fields can be calculated as

$$\mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{A}}{\partial t} \quad , \quad \mathbf{B} = \nabla \times \mathbf{A}$$

Let us now write $\Box \mathbf{A} = 0$ more explicitly as

$$\left(\nabla^2 - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right)\mathbf{A} = 0 \tag{8.50}$$

The general solution of this equation is an, in general infinite, superposition of plane waves of the form

$$\mathbf{A} = \boldsymbol{\epsilon}\, e^{i(\mathbf{k}\cdot\mathbf{r} - \omega t)} \tag{8.51}$$

where $\mathbf{r} = (x^1, x^2, x^3)$ and ϵ is a constant vector and $\mathbf{k}$ is a real vector. Inserting this expression into Eq.(8.50) gives

$$(-|\mathbf{k}|^2 + \frac{\omega^2}{c^2})\mathbf{A} = 0$$

which yields the relation $\omega = \pm|\mathbf{k}|c$. The solutions with $+$ or $-$ simply gives waves moving in opposite directions. We take $\omega = \pm|\mathbf{k}|c$. From the gauge condition $\nabla \cdot \mathbf{A}$ we further find that

$$\mathbf{k} \cdot \boldsymbol{\epsilon} = 0$$

which implies that $\mathbf{A}$ is a transverse wave. If we split $\boldsymbol{\epsilon} = \boldsymbol{\epsilon}_1 + i\boldsymbol{\epsilon}_2$ into a real and an imaginary part this implies that $\mathbf{k}\cdot\boldsymbol{\epsilon}_1 = 0$ as well as $\mathbf{k}\cdot\boldsymbol{\epsilon}_2 = 0$. Our solution is a complex function but since Eq.(8.50) is linear we can find a real solution by taking the real and the imaginary part. To obtain a real solution $\mathbf{A}_r$ we take

$$\mathbf{A}_r = \frac{1}{2}(\mathbf{A} + \mathbf{A}^*) = \boldsymbol{\epsilon}_1 \cos(\mathbf{k}\cdot\mathbf{r} - \omega t) - \boldsymbol{\epsilon}_2 \sin(\mathbf{k}\cdot\mathbf{r} - \omega t)$$

The electric and magnetic fields are then given by

$$\mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{A}_r}{\partial t} = -\frac{\omega}{c}\left[\boldsymbol{\epsilon}_1 \sin(\mathbf{k}\cdot\mathbf{r} - \omega t) + \boldsymbol{\epsilon}_2 \cos(\mathbf{k}\cdot\mathbf{r} - \omega t)\right]$$

$$\mathbf{B} = \nabla \times \mathbf{A}_r = -(\mathbf{k}\times\boldsymbol{\epsilon}_1)\sin(\mathbf{k}\cdot\mathbf{r} - \omega t) - (\mathbf{k}\times\boldsymbol{\epsilon}_2)\cos(\mathbf{k}\cdot\mathbf{r} - \omega t) = \frac{c}{\omega}\mathbf{k}\times\mathbf{E}$$

We can readily check that

$$\mathbf{E} \cdot \mathbf{B} = 0 \quad , \quad \mathbf{E}^2 - \mathbf{B}^2 = 0$$

which are Lorentz invariants. If we chosen $\mathbf{k}$ along the x-axis then the vectors $\boldsymbol{\epsilon}_1$ and $\boldsymbol{\epsilon}_2$ lie in the $y - z$-plane. Let us, for simplicity, take $\boldsymbol{\epsilon}_1$ and $\boldsymbol{\epsilon}_2$ to be orthogonal such that $\boldsymbol{\epsilon}_1 \cdot \boldsymbol{\epsilon}_2 = 0$. Then we can choose the z-axis along $\boldsymbol{\epsilon}_1$ and the y-axis along $\boldsymbol{\epsilon}_2$. Then we can write the electric field as

$$\mathbf{E} = \begin{pmatrix} 0 \\ a\cos(kx - \omega t) \\ b\sin(kx - \omega t) \end{pmatrix}$$

where $k = |\mathbf{k}| = \omega/c$ and

$$\mathbf{B} = \hat{\mathbf{k}} \times \mathbf{E} = \begin{pmatrix} 0 \\ -b\sin(kx - \omega t) \\ a\cos(kx - \omega t) \end{pmatrix}$$

where we used that $\omega\mathbf{k}/c = \hat{\mathbf{k}}$ where $\hat{\mathbf{k}}$ is a unit vector in the $\mathbf{k}$-direction. If we take a fixed value of x then in time the electric field rotates in the $y - z$ plane such that

$$\frac{E_y^2}{a^2} + \frac{E_z^2}{b^2} = 1$$

This figure describes an ellipse in the $y - z$ plane with semi-major and minor axes given by a and b (or b and a if $b > a$) as displayed in Fig. 8.2a. We say that the electromagnetic wave is elliptically polarized. The wave propagates along the x-axis with velocity c since $kx - \omega t = k(x - ct)$. The wave-length is $\lambda = 2\pi/k$ and its frequency $\omega = kc$. If we take the special case $a = 0$ in the equations above then we obtain linearly polarized light with the following non-vanishing components of the electric and magnetic field

$$E_z = b\sin(kx - \omega t)$$
$$B_y = -b\sin(kx - \omega t)$$

This wave is displayed in Fig.8.2b.

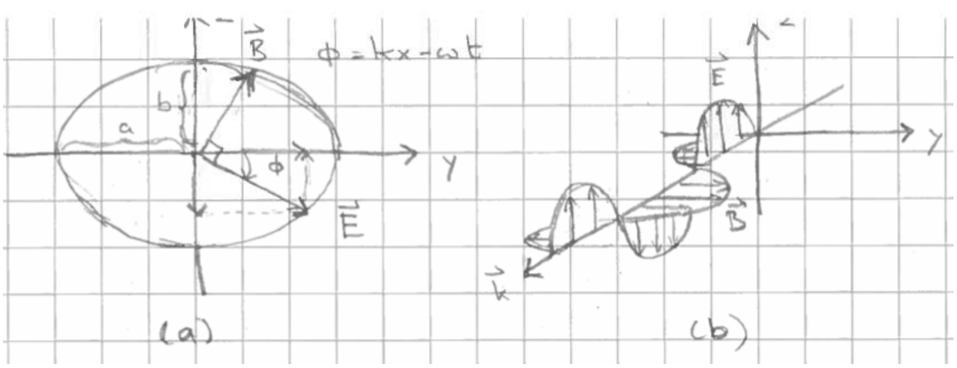

Figure 8.2: An eliptically polarized electromagnetic wave (a) and a linearly polarized wave (b).

If we had not assumed that $\boldsymbol{\epsilon}_1$ and $\boldsymbol{\epsilon}_2$ were orthogonal we would still have found that the light was in general elliptically polarized. The ellipse would simple be rotated in the y_z and a phase shift would appear. We leave it for the reader to show this.

Sofar we considered wave-packets. For any smooth function $\boldsymbol{\epsilon}(\mathbf{k})$ satisfying $\mathbf{k}\cdot\boldsymbol{\epsilon}(\mathbf{k})=0$ (such as $\boldsymbol{\epsilon}(\mathbf{k})=\mathbf{k}\times\mathbf{C}(\mathbf{k})$ with $\mathbf{C}(\mathbf{k})$ any smooth vector) one finds readily that

$$\mathbf{A}(\mathbf{r},t)=\int d\mathbf{k}\,\boldsymbol{\epsilon}(\mathbf{k})\,e^{i(\mathbf{k}\cdot\mathbf{r}-|\mathbf{k}|ct)}$$

is a solution to the wave equation. As above we can split into real and imaginary parts to find real solutions. Now, depending on which boundary conditions we impose, we can find a large variety of solutions. Apart from the plane wave solutions we have found, in particular spherical wave solutions are of common interest as they describe radiation from localized sources.

8.5 The general solution of Maxwell's equations

Let us now see whether we can find a general solution to the equation

$$\Box A^\mu=-\frac{4\pi}{c}j^\mu \tag{8.52}$$

The general solution consists of a homogeneous solution A_h^μ satisfying $\Box A_h^\mu=0$ and a specific solution satisfying $\Box A_s^\mu=-(4\pi/c)j^\mu$. The most general solution is then $A^\mu=A_h^\mu+A_s^\mu$ where a unique solution is provided by specifying initial conditions on the fields. We will assume that charges and currents are localized in space such that $j^\mu\to 0$ when $|x|\to\infty$ and we will therefore look for solutions A^μ that vanish at spatial infinity. If we denote $\mathbf{x}=(x,y,z)$ then Eq.(8.52) has the explicit form

$$\left(-\frac{1}{c^2}\frac{\partial^2}{\partial t^2}+\nabla^2\right)A^\mu(\mathbf{x},t)=-\frac{4\pi}{c}j^\mu(\mathbf{x},t) \tag{8.53}$$

Let us now first write $A(\mathbf{x},t)$ as a Fourier transform in the time variable as

$$A^\mu(\mathbf{x},t)=\int_{-\infty}^{+\infty}d\omega\,A^\mu(\mathbf{x},\omega)\,e^{-i\omega t}$$

then by Fourier transforming Eq.(8.53) we find

$$\left(\frac{\omega^2}{c^2}+\nabla^2\right)A^\mu(\mathbf{x},\omega)=-\frac{4\pi}{c}j^\mu(\mathbf{x},\omega) \tag{8.54}$$

If we can now find a function $D(\mathbf{x}-\mathbf{x}',\omega)$ which satisfies

$$\left(\frac{\omega^2}{c^2}+\nabla^2\right)D(\mathbf{x}-\mathbf{x}',\omega)=-4\pi\delta(\mathbf{x}-\mathbf{x}') \tag{8.55}$$

then a solution to Eq.(8.54) is

$$A^\mu(\mathbf{x},\omega)=\frac{1}{c}\int d\mathbf{x}'\,D(\mathbf{x}-\mathbf{x}',\omega)\,j^\mu(\mathbf{x}',\omega)$$

as follows directly by applying the operator $\omega^2/c^2+\nabla^2$ to both sides of the equation. Since we know that

$$\nabla^2\frac{1}{|\mathbf{x}-\mathbf{x}'|}=-4\pi\delta(\mathbf{x}-\mathbf{x}')$$

we can try a function of the form

$$D(r)=\frac{f(r)}{r}$$

where $r = |\mathbf{x} - \mathbf{x}'|$. A short calculation then shows that

$$\nabla^2 \left(\frac{f(r)}{r} \right) = \frac{1}{r} \frac{\partial^2 f}{\partial r^2} - 4\pi(\mathbf{x} - \mathbf{x}') f(0) \tag{8.56}$$

and we thus find that

$$f(r) = \alpha \, e^{i \frac{\omega}{c} r} + \beta \, e^{-i \frac{\omega}{c} r}$$

where $\alpha + \beta = 1$. This gives

$$D(\mathbf{x} - \mathbf{x}', \omega) = \alpha \, D_R(\mathbf{x} - \mathbf{x}', \omega) + \beta \, D_A(\mathbf{x} - \mathbf{x}', \omega)$$

where

$$D_{R,A}(\mathbf{x} - \mathbf{x}', \omega) = \frac{e^{\pm i \frac{\omega}{c} |\mathbf{x} - \mathbf{x}'|}}{|\mathbf{x} - \mathbf{x}'|}$$

which are called the retarded and advanced propagators. With this expression the solution for A^μ is therefore given by

$$A^\mu(\mathbf{x}, \omega) = \frac{\alpha}{c} \int d\mathbf{x}' \frac{e^{i \frac{\omega}{c} |\mathbf{x} - \mathbf{x}'|}}{|\mathbf{x} - \mathbf{x}'|} \, j^\mu(\mathbf{x}', \omega) + \frac{\beta}{c} \int d\mathbf{x}' \frac{e^{-i \frac{\omega}{c} |\mathbf{x} - \mathbf{x}'|}}{|\mathbf{x} - \mathbf{x}'|} \, j^\mu(\mathbf{x}', \omega)$$

Fourier transforming back to the time domain then gives

$$\begin{aligned} A^\mu(\mathbf{x}, t) &= \int_{-\infty}^{+\infty} d\omega \, A^\mu(\mathbf{x}, \omega) \, e^{-i\omega t} \\ &= \frac{\alpha}{\omega} \int d\mathbf{x}' \int d\omega \frac{j^\mu(\mathbf{x}', \omega)}{|\mathbf{x} - \mathbf{x}'|} \, e^{-i\omega(t - |\mathbf{x} - \mathbf{x}'|/c)} + \frac{\beta}{\omega} \int d\mathbf{x}' \int d\omega \frac{j^\mu(\mathbf{x}', \omega)}{|\mathbf{x} - \mathbf{x}'|} \, e^{-i\omega(t + |\mathbf{x} - \mathbf{x}'|/c)} \\ &= \frac{\alpha}{c} \int d\mathbf{x}' \, \frac{j^\mu(\mathbf{x}', t - |\mathbf{x} - \mathbf{x}'|/c)}{|\mathbf{x} - \mathbf{x}'|} + \frac{\beta}{c} \int d\mathbf{x}' \, \frac{j^\mu(\mathbf{x}', t + |\mathbf{x} - \mathbf{x}'|/c)}{|\mathbf{x} - \mathbf{x}'|} \end{aligned}$$

This is the solution that we were seeking for. We now only need to determine α and β. This is readily done on physical grounds. The first term in the equation depends of the retarded time $t - |\mathbf{x} - \mathbf{x}'|/c$. The contribution to $A^\mu(\mathbf{x}, t)$ comes the current at a point $\mathbf{x}'$ which is a light distance $ct = |\mathbf{x} - \mathbf{x}'|$ in the past. The second term, on the other hand, depends on the advanced time $t + |\mathbf{x} - \mathbf{x}'|/c$ in the future. This contradictions our experience. We therefore need to take $\beta = 0$ and $\alpha = 1$. We therefore find

$$A^\mu(\mathbf{x}, t) = \frac{1}{c} \int d\mathbf{x}' \, \frac{j^\mu(\mathbf{x}', t - |\mathbf{x} - \mathbf{x}'|/c)}{|\mathbf{x} - \mathbf{x}'|} \tag{8.57}$$

This is our final result. We can rewrite this solution a bit in a more Lorentz invariant form. First of all, we can write

$$A^\mu(\mathbf{x}, t) = \frac{1}{c} \int d\mathbf{x}' dt' \, \frac{\delta(t - t' - |\mathbf{x} - \mathbf{x}'|/c)}{|\mathbf{x} - \mathbf{x}'|} \, j^\mu(\mathbf{x}', t') \tag{8.58}$$

Furthermore, since for a function $f(x)$ we have the identity

$$\delta(f(x)) = \sum_j \frac{1}{|\frac{df}{dx}(x_j)|} \delta(x - x_j) \tag{8.59}$$

where x_j are the zeros of f, i.e. $f(x_j) = 0$, it follows that

$$\delta(-c^2(t - t')^2 + (\mathbf{x} - \mathbf{x}')^2) = \frac{1}{2c} \frac{1}{|\mathbf{x} - \mathbf{x}'|} \left(\delta(t - t' - |\mathbf{x} - \mathbf{x}'|/c) + \delta(t - t' + |\mathbf{x} - \mathbf{x}'|/c) \right)$$

where we used the function $f(t') = -c^2(t-t')^2 + (\mathbf{x}-\mathbf{x}')^2$ and

$$\frac{df}{dt'} = 2c^2(t-t') = \pm 2c|\mathbf{x}-\mathbf{x}'| \tag{8.60}$$

If we use the Heaviside function $\theta(x) = 1$ for $x > 0$ and zero otherwise, then we can write

$$\theta(t-t')\delta(-c^2(t-t')^2 + (\mathbf{x}-\mathbf{x}')^2) = \frac{1}{2c}\frac{1}{|\mathbf{x}-\mathbf{x}'|}\delta(t-t'-|\mathbf{x}-\mathbf{x}'|/c)$$

If we then define $D_R(x-x')$ as

$$D_R(x-x') = 2\theta(t-t')\delta(-c^2(t-t')^2 + (\mathbf{x}-\mathbf{x}')^2) = 2\theta(t-t')\,\delta((x-x')^2) \tag{8.61}$$

where $x = (ct, x, y, z) = (x^0, x^1, x^2, x^3)$ and we defined

$$(x-x')^2 = -c^2(t-t')^2 + (\mathbf{x}-\mathbf{x}')^2$$

then we can write Eq.(8.58) as

$$A^\mu(\mathbf{x}, t) = \frac{1}{c}\int d^4x\, D_R(x-x')\, j^\mu(x') \tag{8.62}$$

where d^4x is now a four-dimensional volume element. The function D_R is called a *retarded propagator* and has only contributions from the surface of the past light cone to space-time point x.

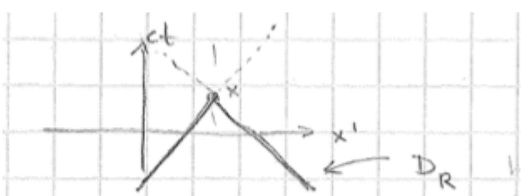

Figure 8.3: The propagator has only values on the past light cone.

This gives a nice description on how the four-potential is determined from initial data.

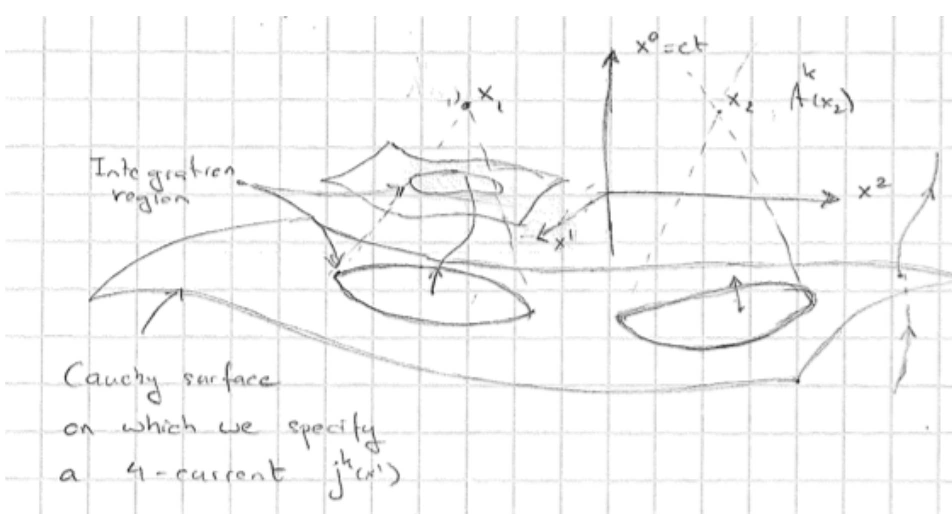

Figure 8.4: Cauchy surface

A Cauchy surface in four-dimensional Minkowski space is a surface with a normal vector that is time-like. If we know the four-current on a Cauchy surface C and know how the charge and current distributions change in time then we can always find the four-potential in space time point to the future of surface C.

We found the general solution Eq.(8.57). However, we did not check that it satisfies the Lorenz gauge condition. This can be seen to be a consequence of charge conservation. Since the integral in Eq.(8.57) goes over all space we can make the substitution $\mathbf{x}' = \mathbf{x} + \mathbf{y}$ and we get

$$A^\mu(\mathbf{x}, t) = \frac{1}{c} \int d\mathbf{y} \, \frac{j^\mu(\mathbf{x} + \mathbf{y}, t - |\mathbf{y}|/c)}{|\mathbf{y}|} \tag{8.63}$$

Denoting $\mathbf{z} = \mathbf{x} + \mathbf{y}$ and $z^0 = ct - |\mathbf{y}|$ we then have

$$\begin{aligned}
\frac{1}{c}\frac{\partial A^0}{\partial t} + \nabla \cdot \mathbf{A} &= \frac{1}{c} \int d\mathbf{y} \, \frac{1}{|\mathbf{y}|} \left(\frac{1}{c}\frac{\partial j^0}{\partial t} + \nabla \cdot \mathbf{j} \right) (\mathbf{x} + \mathbf{y}, t - |\mathbf{y}|/c) \\
&= \frac{1}{c} \int d\mathbf{y} \, \frac{1}{|\mathbf{y}|} \left(\frac{\partial j^0}{\partial z^0} + \nabla_{\mathbf{z}} \cdot \mathbf{j} \right) (\mathbf{z}, z^0)|_{\mathbf{z}=\mathbf{x}+\mathbf{y}, z^0 = t - |\mathbf{y}|/c} = 0
\end{aligned}$$

as a consequence of the fact that $\sum_\nu \partial j^\nu / \partial z^\nu = 0$. So we see that the solution of Eq.(8.57) indeed satisfies the gauge condition.

To the general solution of Eq.(8.57) we can still add arbitrary homogeneous solutions describing the propagation of electromagnetic waves in free space. Such waves are in practice always generated by moving charges (such as antennas far away) and therefore Eq.(8.57) is the more fundamental solution. If we split this solution into a charge and a current part we have

$$\begin{aligned}
\phi(\mathbf{x}, t) &= \int d\mathbf{x}' \, \frac{\rho(\mathbf{x}', t - |\mathbf{x} - \mathbf{x}'|/c)}{|\mathbf{x} - \mathbf{x}'|} \\
\mathbf{A}(\mathbf{x}, t) &= \frac{1}{c} \int d\mathbf{x}' \, \frac{\mathbf{j}(\mathbf{x}', t - |\mathbf{x} - \mathbf{x}'|/c)}{|\mathbf{x} - \mathbf{x}'|}
\end{aligned}$$

We have therefore arrived at the following highlight of this course.

Maxwell's equations:

$$\begin{aligned}
&\nabla \cdot \mathbf{E} = 4\pi \, \rho \\
&-\frac{1}{c}\frac{\partial \mathbf{E}}{\partial t} + \nabla \times \mathbf{B} = \frac{4\pi}{c} \mathbf{j} \\
&\nabla \cdot \mathbf{B} = 0 \\
&\frac{1}{c}\frac{\partial \mathbf{B}}{\partial t} + \nabla \times \mathbf{E} = 0
\end{aligned}$$

Their solutions:

$$\begin{aligned}
\mathbf{E} &= -\nabla \phi - \frac{1}{c}\frac{\partial \mathbf{A}}{\partial t} \\
\mathbf{B} &= \nabla \times \mathbf{A} \\
\phi(\mathbf{x}, t) &= \int d\mathbf{x}' \, \frac{\rho(\mathbf{x}', t - |\mathbf{x} - \mathbf{x}'|/c)}{|\mathbf{x} - \mathbf{x}'|} \\
\mathbf{A}(\mathbf{x}, t) &= \frac{1}{c} \int d\mathbf{x}' \, \frac{\mathbf{j}(\mathbf{x}', t - |\mathbf{x} - \mathbf{x}'|/c)}{|\mathbf{x} - \mathbf{x}'|}
\end{aligned}$$

Chapter 9

Fields from arbitrarily moving point charges

We calculate the four-potential at any space-time point for a a charge in arbitrary motion. The resulting expressions are known as the Liénard-Wiechert potentials which depend on the position and velocities of the particle at the retarded time. From these we calculate the electric and magnetic fields and find that the electric field has a long range component dependent on the acceleration of the particle. Finally we derive Feynman's formula for the field of an accelerated particle and use it to study distant radiation fields.

9.1 The Liénard-Wiechert potentials

In Eq.(8.57) we derived an equation for the four-potential in terms of the four-current. We will rewrite this equation a bit as

$$A^\mu(\mathbf{x},t) = \frac{1}{c}\int d\mathbf{x}'dt'\,\frac{\delta(t-t'-|\mathbf{x}-\mathbf{x}'|/c)}{|\mathbf{x}-\mathbf{x}'|}\,j^\mu(\mathbf{x}',t') \tag{9.1}$$

We will apply this formula to derive the fields of arbitrarily moving point charges. where we consider standard Lorentz frames. Then the current j^μ describes the motion of a point charge with charge q and we have

$$j^0(\mathbf{x},t) = c\rho(\mathbf{x},t) = c\,q\,\delta(\mathbf{x}-\mathbf{z}(t))$$
$$\mathbf{j}(\mathbf{x},t) = q\,\frac{d\mathbf{z}}{dt}(t)\,\delta(\mathbf{x}-\mathbf{z}(t))$$

where $(ct,\mathbf{z}(t))$ describes the world line of the particle in some Lorentz frame. More compactly we can write this in four-vector notation as

$$j^\mu(\mathbf{x},t) = q\,\frac{dz^\mu}{dt}\,\delta(\mathbf{x}-\mathbf{z}(t)) \tag{9.2}$$

where we defined $z^0 = ct$. If we insert this expression into Eq.(8.57) we obtain

$$\begin{aligned} A^\mu(\mathbf{x},t) &= \frac{q}{c}\int d\mathbf{x}'dt'\,\frac{\delta(t-t'-|\mathbf{x}-\mathbf{x}'|/c)}{|\mathbf{x}-\mathbf{x}'|}\,\frac{dz^\mu}{dt'}\,\delta(\mathbf{x}'-\mathbf{z}(t')) \\ &= \frac{q}{c}\int dt'\,\frac{dz^\mu}{dt'}\,\frac{\delta(t-t'-|\mathbf{x}-\mathbf{z}(t')|/c)}{|\mathbf{x}-\mathbf{z}(t')|} \end{aligned} \tag{9.3}$$

We see that the variable t' appears twice inside the delta function. To deal with this situation we use the formula

$$\delta(f(t')) = \sum_i \delta(t'-t_i)\frac{1}{|\frac{df}{dt'}(t_i)|} \quad , \quad f(t_i) = 0 \tag{9.4}$$

Then with

$$f(t') = t - t' - \frac{|\mathbf{x}-\mathbf{z}(t')|}{c}$$

we have

$$\frac{df}{dt'} = -1 + \frac{1}{c}\frac{d\mathbf{z}}{dt'}\cdot\frac{\mathbf{x}-\mathbf{z}(t')}{|\mathbf{x}-\mathbf{z}(t')|} < 0$$

since $|d\mathbf{z}/dt'| < 0$. We these equations we then find from Eq.(9.3) that

$$\begin{aligned} A^\mu(\mathbf{x},t) &= \frac{q}{c}\frac{1}{|\mathbf{x}-\mathbf{z}(t')|}\frac{1}{1-\dfrac{1}{c}\dfrac{d\mathbf{z}}{dt'}\cdot\dfrac{\mathbf{x}-\mathbf{z}(t')}{|\mathbf{x}-\mathbf{z}(t')|}}\frac{dz^\mu}{dt'} \\ &= \frac{q}{c}\frac{1}{|\mathbf{x}-\mathbf{z}(t')| - \dfrac{1}{c}\dfrac{d\mathbf{z}}{dt'}\cdot(\mathbf{x}-\mathbf{z}(t'))}\frac{dz^\mu}{dt'} \end{aligned} \tag{9.5}$$

where the variable t' in this equation must be determined from

$$c(t-t') = |\mathbf{x}-\mathbf{z}(t')| \tag{9.6}$$

for a given $\mathbf{x}$ and t. The time-variable t' determined in this way will be called the *retarded time*. More explicitly, we have from Eq.(9.5) that

$$\phi(\mathbf{x},t) = \frac{q}{|\mathbf{x}-\mathbf{z}(t')| - \dfrac{1}{c}\dfrac{d\mathbf{z}}{dt'}\cdot(\mathbf{x}-\mathbf{z}(t'))} \tag{9.7}$$

$$\mathbf{A}(\mathbf{x},t) = \frac{q}{c}\frac{d\mathbf{z}}{dt'}(t')\frac{1}{|\mathbf{x}-\mathbf{z}(t')| - \dfrac{1}{c}\dfrac{d\mathbf{z}}{dt'}\cdot(\mathbf{x}-\mathbf{z}(t'))} \tag{9.8}$$

The potentials in these two equations are called the *Liénard-Wiechert potentials*. They describe the fact that potentials at space-time point $(\mathbf{x},t)$ depend on the position $\mathbf{z}(t')$ and the velocity $d\mathbf{z}'/dt'$ at an earlier retarded time t' since the fields produced by the moving charge need a time $t-t' = |\mathbf{x}-\mathbf{z}(t')|/c$ to reach the point $\mathbf{x}$ at time t. This is displayed graphically in the figure

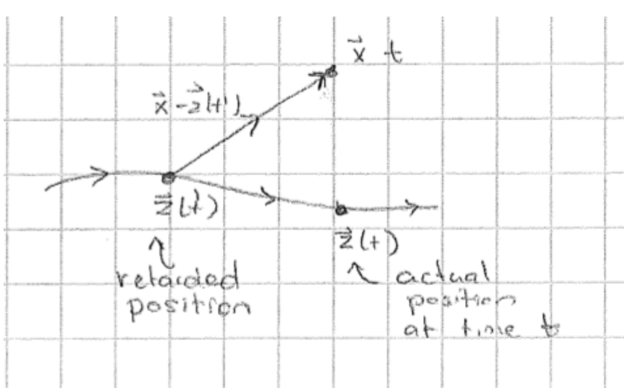

Figure 9.1: Fields at $\mathbf{x}$ at time t originate from the retarded position $\mathbf{z}(t')$ at time $t' = t - |\mathbf{x} - \mathbf{z}(t')|/c$.

As a nice illustration of the Liénard-Wiechert potentials we will show how we can recover Eqs.(8.45) and (8.46) for the uniformly moving charge. Let the charge move with uniform velocity along the x-axis of our Lorentz frame (we use coordinates (ct, x, y, z). Then

$$\mathbf{z}(t') = (vt', 0, 0)$$

From Eqs.(9.7) and (9.8) as well as relation (9.6) we then find

$$\phi(\mathbf{x}, t) = \frac{q}{c(t-t') - \frac{v}{c}(x - vt')} \tag{9.9}$$

$$\mathbf{A}(\mathbf{x}, t) = \frac{q}{c}\frac{q}{c(t-t') - \frac{v}{c}(x - vt')}\begin{pmatrix} v \\ 0 \\ 0 \end{pmatrix} \tag{9.10}$$

It now only remains to solve Eq.(9.6) as a function of t and $\mathbf{x}$. Squaring this equation gives

$$\begin{aligned}
&c^2(t-t')^2 = (\mathbf{x} - \mathbf{z}(t'))^2 = (x - vt')^2 + y^2 + z^2 \quad \Longrightarrow \\
&c^2t^2 - 2c^2tt' + c^2t'^2 = x^2 + y^2 + z^2 - 2xvt' + v^2t'^2 \quad \Longrightarrow \\
&(v^2 - c^2)t'^2 - 2(xv - c^2t)t' + x^2 + y^2 + z^2 - c^2t^2 = 0
\end{aligned}$$

When we solve this quadratic equation for t' (using $t' < t$) we find

$$(1 - \frac{v^2}{c^2})\, t' = t - \frac{vx}{c^2} - \frac{1}{c}\sqrt{(x - vt)^2 + (1 - \frac{v^2}{c^2})(y^2 + z^2)}$$

This gives the following expression for the denominators in Eqs.(9.9) and (9.10).

$$\begin{aligned}
c(t - t') - \frac{v}{c}(x - vt') &= c\left[t - \frac{vx}{c^2} - (1 - \frac{v^2}{c^2})t'\right] \\
&= \sqrt{(x - vt)^2 + (1 - \frac{v^2}{c^2})(y^2 + z^2)}
\end{aligned}$$

We therefore find that Eqs. (9.9) and (9.10) become

$$\phi(\mathbf{x}, t) = \frac{q\gamma}{\sqrt{\gamma^2(x - vt)^2 + y^2 + z^2}}$$

$$\mathbf{A}(\mathbf{x}, t) = \frac{v}{c}\frac{q\gamma}{\sqrt{\gamma^2(x - vt)^2 + y^2 + z^2}}\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

which are exactly Eqs.(8.45) and (8.46) which we found earlier. We will now proceed to write the Liénart-Wiechert potentials in a more Lorentz covariant form. From Eq.(9.5) and (9.6) we find that

$$\begin{aligned}
A^\mu(\mathbf{x}, t) &= \frac{q}{c}\frac{1}{c(t - t') - \dfrac{1}{c}\dfrac{d\mathbf{z}}{dt'}\cdot(\mathbf{x} - \mathbf{z}(t'))}\frac{dz^\mu}{dt'} \\
&= -\frac{q}{-\dfrac{dz^0}{dt'}(x^0 - z^0) + \dfrac{d\mathbf{z}}{dt'}\cdot(\mathbf{x} - \mathbf{z}(t'))}\frac{dz^\mu}{dt'} \\
&= -q\,\frac{dz^\mu}{dt'}\left(\sum_{\nu,\rho} g_{\nu\rho}\frac{dz^\nu}{dt'}(x^\rho - z^\rho(t'))\right)^{-1}
\end{aligned} \tag{9.11}$$

where we used $x^0 = ct$ and $z^0 = ct'$ and $g_{\nu\rho}$ is the metric tensor for the Minkowski metric. If we define

$$x_\nu = \sum_\rho g_{\nu\rho} x^\rho \tag{9.12}$$

then we can write the four-potential as

$$A^\mu(x) = -q\, \frac{dz^\mu}{dt'} \left(\sum_\nu \frac{dz^\nu}{dt'} (x_\nu - z_\nu(t')) \right)^{-1} \tag{9.13}$$

with $x = (x^0, x^1, x^2, x^3)$. We note that in Eq.(9.12) something very strange happens, since the coordinates themselves are regarded as components of a vector. This is only possible since the standard Minkowskian coordinate system has the properties of a vector space. For instance the sum of two coordinates is a another coordinate in the Lorentz frame and we can also multiply with a scalar to obtain another coordinate. This is possible with Cartesian coordinates but, for instance, not for spherical coordinates. The procedure is therefore only well-defined when we restrict ourselves to the standard Minkowskian metric and to Lorentz transformations. This is exactly what we will do in this Chapter.
The four-potential looks much more like a four-vector. However, the presence of the time coordinate t' is still the retarded time in a specific Lorentz frame. Since t' is in one-to-one correspondence with the proper time τ' of the particle at the retarded position we can write

$$\begin{aligned} A^\mu(x) &= -q\, \frac{dz^\mu}{d\tau'} \frac{d\tau'}{dt'} \left(\sum_\nu \frac{dz^\nu}{d\tau'} \frac{d\tau'}{dt'} (x_\nu - z_\nu(t'(\tau'))) \right)^{-1} \\ &= -q\, \frac{dz^\mu}{d\tau'} \frac{1}{\displaystyle\sum_\nu \frac{dz^\nu}{d\tau'} (x_\nu - z_\nu(\tau'))} \end{aligned} \tag{9.14}$$

where with some abuse of notation we denoted $z_\nu(\tau') = z_\nu(t'(\tau'))$. The retarded proper time is then determined from

$$(x^0 - z^0(\tau')) = |\mathbf{x} - \mathbf{z}(\tau')| \tag{9.15}$$

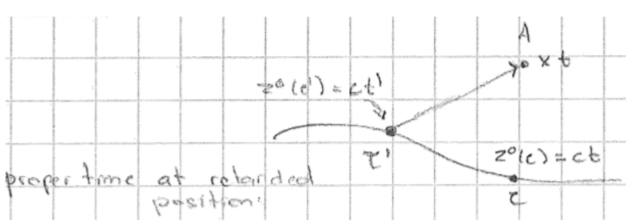

Figure 9.2: Fields at $\mathbf{x}$ at time t originate from the retarded position $\mathbf{z}(\tau')$ at proper time τ'.

From Eq.(9.14) it is now obvious that the four-potential transforms as a four-vector since its numerator is a four-vector and its denominator is a Lorentz invariant. We could also have started directly by directly writing the four-current in terms of the proper time of the particle. We define the four-dimensional delta function

$$\delta^{(4)}(\mathbf{x} - z(\tau)) = \delta(\mathbf{x} - \mathbf{z}(\tau))\, \delta(x^0 - z^0(\tau))$$

where we now parametrized the world line $z(\tau)$ of the particle in terms of its proper time. Then we can write the current as

$$j^\mu(x) = c\, q \int d\tau\, \frac{dz^\mu}{d\tau}\, \delta^{(4)}(x - z(\tau)) \tag{9.16}$$

This is readily shown using Eq.(9.4). We have

$$\begin{aligned} j^\mu(x) &= c\,q \int d\tau\, \frac{dz^\mu}{d\tau}\, \delta(\mathbf{x} - \mathbf{z}(\tau))\, \delta(x^0 - z^0(\tau)) \\ &= c\,q \int d\tau\, \frac{dz^\mu}{d\tau}\, \delta(\mathbf{x} - \mathbf{z}(\tau)) \frac{1}{|\frac{dz^0}{d\tau'}|} \delta(\tau - \tau') = c\,q\, \frac{dz^\mu}{d\tau}\, \frac{1}{|\frac{dz^0}{d\tau}|}\, \delta(\mathbf{x} - \mathbf{z}(\tau)) \end{aligned}$$

Now since $z^0(\tau) = c\,t(\tau)$ we can write

$$j^\mu(x) = q\, \frac{dz^\mu}{dt}\, \frac{dt}{d\tau}\, \frac{1}{\frac{dt}{d\tau}}\, \delta(\mathbf{x} - \mathbf{z}(\tau)) = q\, \frac{dz^\mu}{dt} \delta(\mathbf{x} - \mathbf{z}(t))$$

where we again wrote $\mathbf{z}(t) = \mathbf{z}(t(\tau))$. We see that we exactly recovered Eq. (9.2). We can now insert Eq.(9.16) directly into Eq.(8.62) to give

$$\begin{aligned} A^\mu(x) &= \frac{1}{c} \int d^4x\, D_R(x - x')\, j^\mu(x') \\ &= \frac{2}{c} \int d^4x\, \theta(x^0 - x^{0\prime})\, \delta(\sum_\nu (x - x')_\nu (x - x')^\nu)\, j^\mu(x') \\ &= 2q \int d\tau \int d^4x\, \theta(x^0 - x^{0\prime})\, \delta(\sum_\nu (x - x')_\nu (x - x')^\nu) \frac{dz^\mu}{d\tau}\, \delta^{(4)}(x - z(\tau)) \\ &= 2q \int d\tau\, \theta(x^0 - z^0(\tau))\, \delta(\sum_\nu (x - z(\tau))_\nu (x - z(\tau))^\nu)\, \frac{dz^\mu}{d\tau} \end{aligned}$$

If we now again use relation (9.4) for the delta function we can write this as

$$A^\mu(x) = 2q\, \frac{dz^\mu}{d\tau'}\, \frac{-1}{2 \sum_\nu \frac{dz^\nu}{d\tau'} (x_\nu - z_\nu(\tau'))} = -q \frac{dz^\mu}{d\tau'}\, \frac{1}{\sum_\nu \frac{dz^\nu}{d\tau'} (x_\nu - z_\nu(\tau'))}$$

where the retarded proper time τ' must be determined from the condition

$$0 = \sum_\nu (x_\nu - z_\nu(\tau'))(x^\nu - z^\nu(\tau')) \tag{9.17}$$

where $x^0 > z^0$. This condition is equivalent to Eq.(9.2). We therefore have once again obtained the Liénard-Wiechert potentials.

9.2 The electric and magnetic fields

Having obtained the Liénard-Wiechert potentials it remains to calculate the corresponding electric and magnetic fields. Since we have an explicit formula for A^μ it turns out to be most convenient to calculate the electromagnetic field tensor directly from

$$F_{\mu\nu} = \frac{\partial A_\nu}{\partial x^\mu} - \frac{\partial A_\mu}{\partial x^\nu}$$

where by lowering the index in A^μ we can write

$$A_\nu = -\frac{q}{w} \frac{dz_\nu}{d\tau'} \tag{9.18}$$

where for compactness of notation we further defined

$$w = \sum_\nu \frac{dz^\nu}{d\tau'}(x_\nu - z_\nu(\tau'))$$

Let us first calculate $\partial A_\nu/\partial x^\mu$. We have

$$\frac{\partial A_\nu}{\partial x^\mu} = -q\left[\frac{d^2 z_\nu}{d\tau'^2}\frac{\partial \tau'}{\partial x^\mu}\frac{1}{w} - \frac{dz_\nu}{d\tau'}\frac{\partial w}{\partial x^\mu}\frac{1}{w^2}\right] \tag{9.19}$$

The calculation is somewhat complicated due to the implicit dependence of τ' on x^μ through the condition (9.17). The derivative $\partial\tau'/\partial x^\mu$ can, however, be obtained by differentiating Eq.(9.17) with respect to x^μ. This gives

$$\begin{aligned}
0 &= \frac{\partial}{\partial x^\mu}\left[\sum_{\nu,\rho} g_{\nu\rho}(x^\nu - z^\nu(\tau'))(x^\rho - z^\rho(\tau'))\right] = 2\sum_\nu (x_\nu - z_\nu(\tau'))\frac{\partial}{\partial x^\mu}(x^\nu - z^\nu(\tau')) \\
&= 2\sum_\nu (x_\nu - z_\nu(\tau'))\left[\delta_{\mu\nu} - \frac{dz^\nu}{d\tau'}\frac{\partial\tau'}{\partial x^\mu}\right] = 2(x_\mu - z_\mu(\tau')) - 2\frac{\partial \tau'}{\partial x^\mu}\sum_\nu (x_\nu - z_\nu(\tau'))\frac{dz^\nu}{d\tau'} \\
&= 2(x_\mu - z_\mu(\tau')) - 2\,w\,\frac{\partial\tau'}{\partial x^\mu}
\end{aligned}$$

We therefore find that

$$\frac{\partial \tau'}{\partial x^\mu} = \frac{1}{w}(x_\mu - z_\mu(\tau')) = \frac{y_\mu}{w} \tag{9.20}$$

where we further defined $y_\mu = x_\mu - z_\mu(\tau')$. We further see that we need to evaluate $\partial w/\partial x^\mu$ in Eq.(9.19). We find

$$\begin{aligned}
\frac{\partial w}{\partial x^\mu} &= \frac{\partial}{\partial x^\mu}\left(\sum_\nu \frac{dz^\nu}{d\tau'}(x_\nu - z_\nu(\tau'))\right) \\
&= \sum_\nu \left[\frac{d^2 z_\nu}{d\tau'}\frac{\partial\tau'}{\partial x^\mu}(x^\nu - z^\nu) + \frac{dz_\nu}{d\tau'}\left(\delta_{\mu\nu} - \frac{dz^\nu}{d\tau'}\frac{\partial \tau'}{\partial x^\mu}\right)\right] \\
&= \frac{dz_\mu}{d\tau'} + \frac{\partial \tau'}{\partial x^\mu}\sum_\nu\left(\frac{d^2 z_\nu}{d\tau'}(x^\nu - z^\nu) - \frac{dz^\nu}{d\tau'}\frac{dz_\nu}{d\tau'}\right) = \frac{dz_\mu}{d\tau'} + \frac{\partial\tau'}{\partial x^\mu}\frac{dw}{d\tau'}
\end{aligned}$$

If we insert this result into Eq.(9.19) we find

$$\begin{aligned}
\frac{\partial A_\nu}{\partial x^\mu} &= -q\left[\frac{d^2 z_\nu}{d\tau'^2}\frac{\partial\tau'}{\partial x^\mu}\frac{1}{w} - \frac{dz_\nu}{d\tau'}\frac{1}{w^2}\left(\frac{dz_\mu}{d\tau'} + \frac{\partial \tau'}{\partial x^\mu}\frac{dw}{d\tau'}\right)\right] \\
&= -\frac{q}{w^2}\left[\left(w\,\frac{d^2 z_\nu}{d\tau'^2} - \frac{dw}{d\tau'}\frac{dz_\nu}{d\tau'}\right)\frac{\partial\tau'}{\partial x^\mu} - \frac{dz_\nu}{d\tau'}\frac{dz_\mu}{d\tau'}\right]
\end{aligned}$$

We can now insert expression (9.20) and use $dz_\mu/d\tau' = -dy_\mu/d\tau'$ to obtain

$$\begin{aligned}
\frac{\partial A_\nu}{\partial x^\mu} &= -\frac{q}{w^2}\left[\left(w\,\frac{d^2 z_\nu}{d\tau'^2} - \frac{dw}{d\tau'}\frac{dz_\nu}{d\tau'}\right)\frac{y_\mu}{w} + \frac{dz_\nu}{d\tau'}\frac{dy_\mu}{d\tau'}\right] \\
&= -\frac{q}{w}\left[\frac{1}{w}\frac{d^2 z_\nu}{d\tau'^2} - \frac{1}{w^2}\frac{dw}{d\tau'}\frac{dz_\nu}{d\tau'}\,y_\mu + \frac{1}{w}\frac{dz_\nu}{d\tau'}\frac{dy_\mu}{d\tau'}\right] = -\frac{q}{w}\frac{d}{d\tau'}\left[\frac{1}{w}\frac{dz_\nu}{d\tau'}\,y_\mu\right]
\end{aligned}$$

From this equation we therefore find the relatively compact formula

$$F_{\mu\nu} = \frac{q}{w}\frac{d}{d\tau'}\left[\frac{1}{w}\left(\frac{dz_\mu}{d\tau'}\,y_\nu - \frac{dz_\nu}{d\tau'}\,y_\mu\right)\right] \tag{9.21}$$

We see that this expression correctly transforms as a tensor under Lorentz transformations since τ' and w are Lorentz invariants. The form of this expression is invariant under re-parametrization of the path of the particle. If we have a new path parameter $s(\tau')$ then

$$w = \sum_\nu \frac{dz^\nu}{d\tau'}(x_\nu - z_\nu(\tau')) = \sum_\nu \frac{dz^\nu}{ds}\frac{ds}{d\tau'}(x_\nu - z_\nu(\tau')) = u\,\frac{ds}{d\tau'}$$

where

$$u = \sum_\nu \frac{dz^\nu}{ds}(x_\nu - z_\nu(s)) \tag{9.22}$$

and we denote $z_\nu(s(\tau)) = z_\nu(\tau')$. Since $d/d\tau' = ds/d\tau' d/ds$ it then follows that

$$F_{\mu\nu} = \frac{q}{u}\frac{d}{ds}\left[\frac{1}{u}\left(\frac{dz_\mu}{ds}\,y_\nu - \frac{dz_\nu}{ds}\,y_\mu\right)\right] \tag{9.23}$$

where the right hand side is evaluated at value of s which satisfies the condition

$$0 = \sum_\nu (x_\nu - z_\nu(s))(x^\nu - z^\nu(s)) \tag{9.24}$$

Let us work our Eq.(9.25) a bit further. If we denote the first and second derivative of z_ν with respect to s as $\dot z_\nu$ and $\ddot z_\nu$ we have

$$\begin{aligned} F_{\mu\nu} &= \frac{q}{u}\left[-\frac{\dot u}{u^2}(\dot z_\mu y_\nu - \dot z_\nu y_\mu) + \frac{1}{u}(\ddot z_\mu y_\nu - \ddot z_\nu y_\mu + \dot z_\mu \dot y_\nu - \dot z_\nu \dot y_\mu)\right] \\ &= \frac{q}{u^3}\left[\,u\,(\ddot z_\mu y_\nu - \ddot z_\nu y_\mu) - \dot u\,(\dot z_\mu y_\nu - \dot z_\nu y_\mu)\,\right] \end{aligned} \tag{9.25}$$

where in the second step we used that $\dot z_\nu = -\dot y_\nu$. We are now quite close to the end of our calculation. We parametrize the path of the particle in a given Lorentz frame as

$$z^\mu(t') = (ct', \mathbf{z}(t'))$$

i.e. we choose $s = t'$ and hence $\dot z_\nu = dz_\nu/dt'$ etc.. Then by lowering the index we have

$$z_\mu(t') = (-ct', \mathbf{z}(t'))$$

The electric field is given by $E_j = -F_{0j}$ for $j = 1, 2, 3$ and therefore

$$\begin{aligned} E_j &= -\frac{q}{u^3}\left[\,u\,(\ddot z_0 y_j - \ddot z_j y_0) - \dot u\,(\dot z_0 y_j - \dot z_j y_0)\,\right] \\ &= -\frac{q}{u^3}\left[u(c(t-t'))\ddot z_j - \dot u(-c(x_j - z_j) - \dot z_j(-c(t-t')))\right] \end{aligned} \tag{9.26}$$

Let us write this in vector notation. If we further use that $c(t-t') = |\mathbf{x} - \mathbf{z}(t')|$ we can write

$$\mathbf{E} = -\frac{q}{u^3}\left[\dot u\,(-\dot{\mathbf{z}}\,|\mathbf{x}-\mathbf{z}| + c\,(\mathbf{x}-\mathbf{z})) + u\,\ddot{\mathbf{z}}\,|\mathbf{x}-\mathbf{z}|\right] \tag{9.27}$$

It only remains to evaluate u and $\dot u$. We have

$$\begin{aligned} u &= \sum_\nu \frac{dz^\nu}{dt'}(x_\nu - z_\nu(t')) = -c^2(t-t') + \frac{d\mathbf{z}}{dt'}\cdot(\mathbf{x}-\mathbf{z}) \\ &= -c|\mathbf{x}-\mathbf{z}| + \frac{d\mathbf{z}}{dt'}\cdot(\mathbf{x}-\mathbf{z}) = -c\,|\mathbf{x}-\mathbf{z}|\left(1 - \frac{1}{c}\,\mathbf{n}\cdot\dot{\mathbf{z}}\right) \end{aligned} \tag{9.28}$$

where we defined the normal vector

$$\mathbf{n} = \frac{\mathbf{x}-\mathbf{z}}{|\mathbf{x}-\mathbf{z}|}$$

Further

$$\dot{u} = \sum_{\nu} \left[\frac{d^2 z^{\nu}}{dt'^2}(x_{\nu} - z_{\nu}(t')) - \frac{dz^{\nu}}{dt'}\frac{dz_{\nu}}{dt'} \right] \tag{9.29}$$

where

$$\dot{z} = (c, \dot{\mathbf{z}}) \quad , \quad \ddot{z} = (0, \ddot{\mathbf{z}})$$

This gives

$$\dot{u} = \ddot{\mathbf{z}} \cdot (\mathbf{x}-\mathbf{z}) - |\dot{\mathbf{z}}|^2 + c^2$$

Collecting our results we then find from Eq.(9.27) that

$$\begin{aligned}
\mathbf{E} &= \frac{q}{c^3|\mathbf{x}-\mathbf{z}|^3 \left(1 - \frac{1}{c}\,\mathbf{n}\cdot\dot{\mathbf{z}}\right)^3} \\
&\quad \times \left\{ -c|\mathbf{x}-\mathbf{z}|^2\ddot{\mathbf{z}} \left(1 - \frac{1}{c}\,\mathbf{n}\cdot\dot{\mathbf{z}}\right) + [\ddot{\mathbf{z}}\cdot(\mathbf{x}-\mathbf{z}) + c^2 - |\dot{\mathbf{z}}|^2][-\dot{\mathbf{z}}\,|\mathbf{x}-\mathbf{z}| + c\,(\mathbf{x}-\mathbf{z})] \right\} \\
&= \frac{q}{|\mathbf{x}-\mathbf{z}| \left(1 - \frac{1}{c}\,\mathbf{n}\cdot\dot{\mathbf{z}}\right)^3} \\
&\quad \times \left\{ -\frac{\ddot{\mathbf{z}}}{c^2}\left(1 - \frac{1}{c}\,\mathbf{n}\cdot\dot{\mathbf{z}}\right) + \frac{1}{c^2}\,\ddot{\mathbf{z}}\cdot\mathbf{n}\left(\mathbf{n} - \frac{\dot{\mathbf{z}}}{c}\right) + \left(1 - \frac{|\dot{\mathbf{z}}|^2}{c^2}\right)\frac{1}{|\mathbf{x}-\mathbf{z}|}\left(\mathbf{n} - \frac{\dot{\mathbf{z}}}{c}\right) \right\}
\end{aligned} \tag{9.30}$$

Below we will show how to write this expression in a more convenient form. Let us, however, first calculate the magnetic field. We have

$$\mathbf{B} = (F_{23}, -F_{13}, F_{12})$$

We then have from Eq.(9.25), for instance, that

$$B_1 = F_{23} = \frac{q}{u^3}\left(u\,(\ddot{z}_2 y_3 - \ddot{z}_3 y_2) - \dot{u}\,(\dot{z}_2 y_3 - \dot{z}_3 y_2)\right)$$

and similarly for the other components for the $\mathbf{B}$ field. We can write this in vector notation as

$$\begin{aligned}
\mathbf{B} &= \frac{q}{u^3}\left(u\,\ddot{\mathbf{z}} \times (\mathbf{x}-\mathbf{z}) - \dot{u}\,\dot{\mathbf{z}} \times (\mathbf{x}-\mathbf{z})\right) \\
&= \frac{q}{u^3}\left(u\,\ddot{\mathbf{z}}\,|\mathbf{x}-\mathbf{z}| - \dot{u}\,\dot{\mathbf{z}}\,|\mathbf{x}-\mathbf{z}|\right) \times \frac{\mathbf{x}-\mathbf{z}}{|\mathbf{x}-\mathbf{z}|}
\end{aligned}$$

Comparison with Eq.(9.27) then gives

$$\mathbf{B} = -\mathbf{E} \times \frac{\mathbf{x}-\mathbf{z}}{|\mathbf{x}-\mathbf{z}|} = \mathbf{n} \times \mathbf{E} \tag{9.31}$$

The magnetic field can therefore be simply calculated when the electric field is known. Equations (9.30) and (9.31) are the main results of our derivation. We will now proceed to simplify expression (9.30). Let us start by defining the quantities

$$\begin{aligned}
R &= |\mathbf{x} - \mathbf{z}(t')| \\
\kappa &= 1 - \frac{1}{c}\,\mathbf{n}(t')\cdot\dot{\mathbf{z}}(t') \\
\boldsymbol{\beta} &= \frac{1}{c}\,\dot{\mathbf{z}}(t')
\end{aligned}$$

Then we can write (9.30) as

$$\mathbf{E} = \frac{q}{\kappa^3 R}\left\{-\frac{1}{c}\dot{\boldsymbol{\beta}}\,\kappa + \frac{1}{c}\dot{\boldsymbol{\beta}}\cdot\mathbf{n}\,(\mathbf{n}-\boldsymbol{\beta}) + (1-|\boldsymbol{\beta}|^2)\frac{1}{R}(\mathbf{n}-\boldsymbol{\beta})\right\} \tag{9.32}$$

If we use that $\mathbf{a}\times(\mathbf{b}\times\mathbf{c}) = \mathbf{b}\,(\mathbf{a}\cdot\mathbf{c}) - \mathbf{c}\,(\mathbf{a}\cdot\mathbf{b})$ then we see that

$$\mathbf{n}\times[(\mathbf{n}-\boldsymbol{\beta})\times\dot{\boldsymbol{\beta}}] = (\mathbf{n}-\boldsymbol{\beta})\,\dot{\boldsymbol{\beta}}\cdot\mathbf{n} - \dot{\boldsymbol{\beta}}\,\mathbf{n}\cdot(\mathbf{n}-\boldsymbol{\beta}) = (\mathbf{n}-\boldsymbol{\beta})\,\dot{\boldsymbol{\beta}}\cdot\mathbf{n} - \kappa\,\dot{\boldsymbol{\beta}}$$

and we therefore see that we can write

$$\begin{aligned}\mathbf{E} &= \frac{q}{\kappa^3 R}\left\{\frac{1}{c}\,\mathbf{n}\times[(\mathbf{n}-\boldsymbol{\beta})\times\dot{\boldsymbol{\beta}}] + (1-|\boldsymbol{\beta}|^2)\frac{1}{R}(\mathbf{n}-\boldsymbol{\beta})\right\}\\ &= \frac{q}{\kappa^3 R^2}\,(1-|\boldsymbol{\beta}|^2)(\mathbf{n}-\boldsymbol{\beta}) + \frac{q}{c\,\kappa^3 R}\,\mathbf{n}\times[(\mathbf{n}-\boldsymbol{\beta})\times\dot{\boldsymbol{\beta}}]\end{aligned} \tag{9.33}$$

We remind the reader that all the quantities on the right hand side depend on the retarded time t' that must be obtained by solved Eq.(9.6). The electric field therefore consists of a term that decays as $1/R^2$ which depends on the velocity $\boldsymbol{\beta}$ and term that decays as $1/R$ which depends both on the velocity and the acceleration. The important point to stress here is that accelerated charges are a source of fields that fall off slowly. It is therefore exactly the second term in Eq.(9.33) that makes long range communications using light or radio signals possible.

9.3 Feynman's formula

The main results of the previous Section were Eqs.(9.33) and (9.31). They allow us to calculate the electric and magnetic fields from an arbitrarily moving charge provided we can solve Eq.(9.6) for the retarded time t'. Instead of using derivatives with respect to the retarded time we can rewrite Eq.(9.33) in a different form by using derivatives with respect to t instead. In that case it attains the form

$$\mathbf{E} = q\left[\frac{\mathbf{n}}{R^2} + \frac{R}{c}\frac{d}{dt}\left(\frac{\mathbf{n}}{R^2}\right) + \frac{1}{c^2}\frac{d^2}{dt^2}\mathbf{n}\right] \tag{9.34}$$

Since the retarded time t' is a function the space-time point $(\mathbf{x}, t)$ the quantities $\mathbf{n}(t')$ and $R = |\mathbf{x} - \mathbf{z}(t')|$ appearing on the right hand side can be equivalently regarded as functions of $(\mathbf{x}, t)$. Feynman discusses this formula at some length in his famous Feynman Lectures on Physics, so we will simply call it *Feynman's formula*. Let us now proceed by showing that Eq.(9.34) is equivalent to Eq.(9.33). We start by writing Eqs.(9.7) and (9.8) as

$$\phi(\mathbf{x}, t) = \frac{q}{\kappa R} \tag{9.35}$$

$$\mathbf{A}(\mathbf{x}, t) = \frac{q}{c\kappa R}\frac{d\mathbf{z}}{dt'} = -\frac{q}{c\kappa R}\frac{d}{dt'}(\mathbf{x} - \mathbf{z}(t')) = -\frac{q}{c\kappa R}\frac{d}{dt'}\mathbf{R} \tag{9.36}$$

where we defined $\mathbf{R} = \mathbf{x} - \mathbf{z}(t')$. Let us further derive a few useful relations. First of all, by taking the derivative of the condition Eq.(9.6) with respect to t we find

$$c(1 - \frac{\partial t'}{\partial t}) = -\frac{\mathbf{x} - \mathbf{z}}{|\mathbf{x} - \mathbf{z}|}\cdot\frac{d\mathbf{z}}{dt'}\frac{\partial t'}{\partial t}$$

and therefore

$$\frac{\partial t'}{\partial t} = \frac{1}{1 - \frac{1}{c}\mathbf{n}\cdot\dot{\mathbf{z}}} = \frac{1}{\kappa}$$

where $\dot{\mathbf{z}} = d\mathbf{z}/dt'$. We can therefore write

$$\frac{\partial}{\partial t} = \frac{\partial t'}{\partial t}\frac{\partial}{\partial t'} = \frac{1}{\kappa}\frac{\partial}{\partial t'} \tag{9.37}$$

We further have
$$\frac{dR}{dt} = \frac{d}{dt}|\mathbf{x}-\mathbf{z}(t')| = -\frac{\mathbf{x}-\mathbf{z}}{|\mathbf{x}-\mathbf{z}|}\cdot\frac{d\mathbf{z}}{dt'}\frac{\partial t'}{\partial t} = -\frac{1}{\kappa}\mathbf{n}\cdot\dot{\mathbf{z}}$$

and hence
$$1 - \frac{1}{c}\frac{dR}{dt} = 1 + \frac{1}{\kappa c}\mathbf{n}\cdot\dot{\mathbf{z}} = \frac{1}{\kappa}(\kappa + \frac{1}{c}\mathbf{n}\cdot\dot{\mathbf{z}}) = \frac{1}{\kappa}$$

Using this relation and Eq.(9.37) we can rewrite Eqs.(9.35) and (9.36) as

$$\phi(\mathbf{x},t) = \frac{q}{R}\left(1 - \frac{1}{c}\frac{dR}{dt}\right) \tag{9.38}$$

$$\mathbf{A}(\mathbf{x},t) = -\frac{q}{cR}\frac{d}{dt}\mathbf{R} = -\frac{q}{cR}\frac{d}{dt}(\mathbf{n}R) \tag{9.39}$$

These equations are the starting point of our derivation as in terms of these potentials the electric field is given by
$$\mathbf{E} = -\nabla\phi - \frac{1}{c}\frac{\partial\mathbf{A}}{\partial t}$$

Let us start with the first term. From Eq.(9.38) we have

$$-\nabla\phi = \frac{q}{R^2}\nabla R - \frac{q}{c}\frac{\nabla R}{R^2}\frac{dR}{dt} + \frac{q}{c}\frac{1}{R}\frac{d}{dt}\nabla R \tag{9.40}$$

It remains to calculate ∇R. We have

$$\nabla R = \nabla|\mathbf{x}-\mathbf{z}(t')| = \frac{\mathbf{x}-\mathbf{z}}{|\mathbf{x}-\mathbf{z}|} + \nabla t'\frac{d}{dt'}|\mathbf{x}-\mathbf{z}(t')| = \mathbf{n} + \frac{dR}{dt'}\nabla t'$$

where we must not forget that t' depends on $\mathbf{x}$ through condition (9.6). If we take the gradient of this condition we find

$$-c\nabla t' = \mathbf{n} - \frac{\mathbf{x}-\mathbf{z}}{|\mathbf{x}-\mathbf{z}|}\cdot\frac{d\mathbf{z}}{dt'}\nabla t' \implies \nabla t' = -\frac{1}{c}\mathbf{n} + \frac{1}{c}\mathbf{n}\cdot\frac{d\mathbf{z}}{dt'}\nabla t'$$
$$\implies \nabla t' = -\frac{\mathbf{n}}{\kappa c}$$

and therefore
$$\nabla R = \mathbf{n} - \frac{\mathbf{n}}{\kappa c}\frac{dR}{dt'} = \mathbf{n}\left(1 - \frac{1}{c}\frac{dR}{dt}\right)$$

If we insert this result into Eq.(9.40) we obtain

$$-\nabla\phi = \frac{q}{R^2}\mathbf{n}\left(1-\frac{1}{c}\frac{dR}{dt}\right) - \frac{q}{c}\frac{\mathbf{n}}{R^2}\left(1-\frac{1}{c}\frac{dR}{dt}\right)\frac{dR}{dt} + \frac{q}{cR}\frac{d}{dt}\left[\mathbf{n}\left(1-\frac{1}{c}\frac{dR}{dt}\right)\right]$$

We obtain the electric field if we add to this expression the term $-(1/c)\partial\mathbf{A}/\partial$ using Eq.(9.39). This gives

$$\frac{1}{q}\mathbf{E} = \frac{\mathbf{n}}{R^2} - \frac{1}{c}\frac{2\mathbf{n}}{R^2}\frac{dR}{dt} + \frac{\mathbf{n}}{R^2c^2}\left(\frac{dR}{dt}\right)^2 + \frac{1}{Rc}\frac{d\mathbf{n}}{dt} - \frac{1}{c^2R}\frac{d}{dt}\left(\mathbf{n}\frac{dR}{dt}\right) + \frac{1}{c^2}\frac{d}{dt}\left(\frac{1}{R}\frac{d}{dt}[R\mathbf{n}]\right)$$
$$= \frac{\mathbf{n}}{R^2} + \frac{R}{c}\frac{d}{dt}\left(\frac{\mathbf{n}}{R^2}\right) + \frac{1}{c^2}\frac{d^2}{dt^2}\mathbf{n}$$

This gives exactly the equation (9.34) that we wanted to prove. Eqs.(9.33) or (9.34) together with the corresponding magnetic field of Eq.(9.31) are fundamental equations in the theory of electromagnetism. Both Eqs.(9.33) or (9.34) have their own advantages in doing calculations. We will dicuss the Eq.(9.33) in more detail later when we discuss radiation of moving charges. Let us therefore discuss Eq.(9.34). The first two terms in Eq.(9.34) look very much the first two terms of a Taylor expansion of $\mathbf{n}(t)/R^2(t)$ around the retarded time $t - R/c$, i.e.

$$\frac{\mathbf{n}(t)}{R^2(t)} = \frac{\mathbf{n}(t - R/c)}{R^2(t - R/c)} + \frac{R}{c}\frac{d}{dt}\left(\frac{\mathbf{n}(t - R/c)}{R^2(t - R/c)}\right) + O((R/c)^2)$$

Therefore, up to terms of order $(R/c)^2$ the first two terms of Eq.(9.34) yield the same result as that of an instantaneous electric field of a point charge at position $\mathbf{x} - \mathbf{z}(t)$. The last term, proportional to $d^2\mathbf{n}/dt^2$ is the important one for emitted radiation as it decays slowly as $1/R$ as we will see. Let us analyze this term in an example.

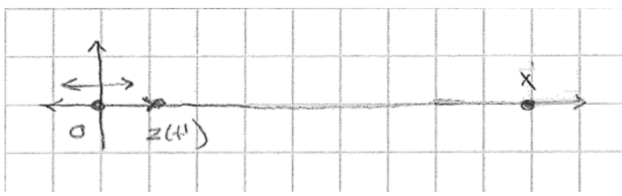

Figure 9.3: Fields at $\mathbf{x}$ due to a charge oscillation in the horizontal direction.

We imagine a charge q oscillating round the origin O and we look at the charge from a long distance (point $\mathbf{x}$ in the figure). In the first case the charge is moving back and forth towards us. In this case

$$\mathbf{n}(t') = \frac{\mathbf{x} - \mathbf{z}(t')}{|\mathbf{x} - \mathbf{z}(t')|} = 1$$

and hence $d^2\mathbf{n}/dt^2 = 0$. We do not see any radiation in this case. In the next case we let the charge move up and down.

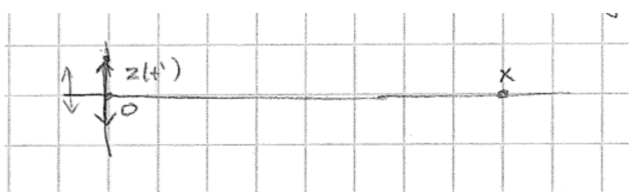

Figure 9.4: Fields at $\mathbf{x}$ due to a charge oscillation in the vertical direction.

In this case, since $|\mathbf{x} - \mathbf{z}(t')| \approx |\mathbf{x}|$ we have

$$\frac{d^2}{dt^2}\mathbf{n} \approx \frac{d^2}{dt^2}\frac{\mathbf{x} - \mathbf{z}(t - |\mathbf{x}|/c)}{|\mathbf{x}|} = -\frac{1}{|\mathbf{x}|}\frac{d^2}{dt^2}\mathbf{z}(t - \frac{|\mathbf{x}|}{c})$$

So in this case the electric field in $\mathbf{x}$ is approximately

$$\mathbf{E} = -\frac{q}{|\mathbf{x}|c^2}\frac{d^2}{dt^2}\mathbf{z}(t - \frac{|\mathbf{x}|}{c})$$

whereas the corresponding magnetic field is given by Eq.(9.31) as

$$\mathbf{B} = \mathbf{n} \times \mathbf{E} = -\frac{q}{|\mathbf{x}|c^2}\mathbf{n} \times \frac{d^2}{dt^2}\mathbf{z}(t - \frac{|\mathbf{x}|}{c})$$

We have now approximately determined the radiation field in two cases. The next thing to do, obviously, would be to find the angular distribution of the radiation. We will approach this problem a bit more systematically later.

Chapter 10

The energy-momentum tensor

We derive the energy and momentum conservation laws of the electromagnetic field in differential and integral form and show that these can elegantly written as the divergence of a symmetric energy-momentum tensor. We further derive a corresponding tensor for the charged particles and describe the energy and momentum conservation laws in general coordinates for the complete field and matter system.

10.1 Conservation of energy and momentum of fields and particles

In the previous Chapter we saw that an oscillating charge produces radiation. This radiation, via the Lorentz force, can set other charges in motion for away which is in fact the principle behind the antenna. This implies that both energy and momentum is transported from one charge to another charges. This in turn means that that energy and momentum must be transported by the electromagnetic field. We have to figure out how this happens.
Imagine two particles interacting via each other's electromagnetic field.

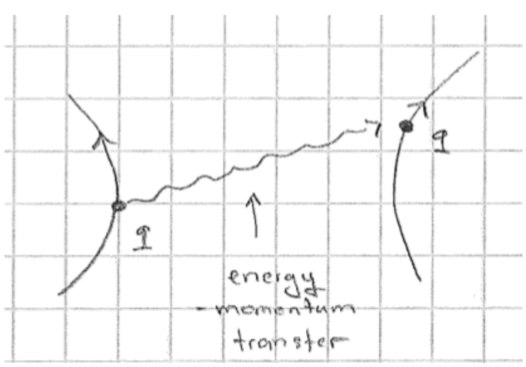

Figure 10.1: Energy and momentum transfer between two moving charges.

Since the energy and momentum of each particle can change there will be a transfer of energy and momentum by the field. The total energy E and momentum $\mathbf{P}$ of the system must be constant, so we have

$$E_{\text{field}} + E_{\text{particles}} = E = \text{constant}$$
$$\mathbf{P}_{\text{field}} + \mathbf{P}_{\text{particles}} = \mathbf{P} = \text{constant}$$

We know that the energy and momentum of the particles is given by the Lorentz force law. It then remains to find the energy and momentum of the field. Let us start by discussing the energy. The change in energy E_{p} of a particle was given by Eq. (6.40) as

$$\frac{dE_{\text{p}}}{dt} = \frac{d}{dt}(\gamma\, mc^2) = q\,\mathbf{E}\cdot\mathbf{v}$$

where $\mathbf{v}$ is the velocity of the particle. To relate this expression to Maxwell's equations we write it as an integral over a current

$$\frac{dE_{\text{p}}}{dt} = \int d\mathbf{x}\,\mathbf{E}(\mathbf{x},t)\cdot\mathbf{j}(\mathbf{x},t) \tag{10.1}$$

where

$$\mathbf{j}(\mathbf{x},t) = q\,\frac{d\mathbf{z}(t)}{dt}\,\delta(\mathbf{x}-\mathbf{z}(t))$$

where $\mathbf{z}(t)$ is the path of the particle in the Lorentz frame that we are using. It is clear that we can also do this if we have several particles. If we have N particles with charge q_j and paths $\mathbf{z}_j(t)$ then we can write

$$\mathbf{j}(\mathbf{x},t) = \sum_{j=1}^{N} q_j\,\frac{d\mathbf{z}_j(t)}{dt}\,\delta(\mathbf{x}-\mathbf{z}_j(t)) \tag{10.2}$$

and Eq.(10.1) is still valid. Let us imagine that these particles are moving around in some volume V as in the figure

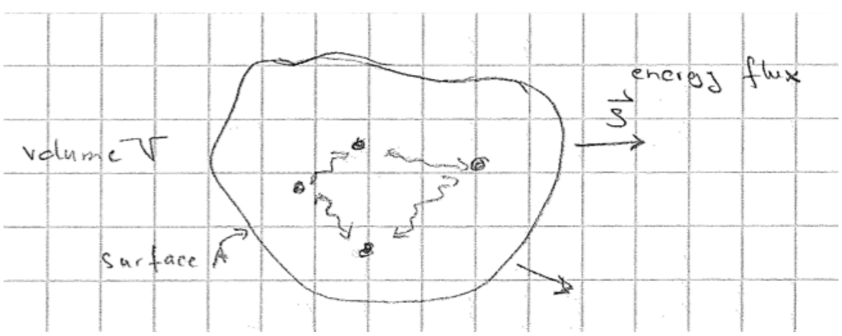

Figure 10.2: Charges moving around in a volume V

The particles interact with each other and radiate. Some of the radiation will pass through the surface A of the volume. We therefore have an outgoing energy flux $\mathbf{S}$ passing the surface. The energy conservation law then tells us that

$$\int_A \mathbf{S}\cdot d\mathbf{A} = -\frac{d}{dt}(E_{\text{p}} + E_{\text{f}}) \tag{10.3}$$

where E_{f} is the energy of the field inside volume V. If we now write the field energy as an integral of an energy density $u(\mathbf{x},t)$ as

$$E_{\text{f}} = \int_V d\mathbf{x}\,u(\mathbf{x},t)$$

then with help of Eq.(10.1) we can write the energy conservation law (10.3) as

$$\int_A \mathbf{S}\cdot d\mathbf{A} = -\int_V d\mathbf{x}\,\mathbf{E}(\mathbf{x},t)\cdot\mathbf{j}(\mathbf{x},t) - \int_V d\mathbf{x}\,\frac{\partial u(\mathbf{x},t)}{\partial t}$$

Then by further writing the flux integral as a divergence

$$\int_A \mathbf{S}\cdot d\mathbf{A} = \int_V d\mathbf{x}\,\nabla\cdot\mathbf{S}$$

we find the relation

$$0 = \int_V d\mathbf{x}\,\left(\frac{\partial u}{\partial t} + \nabla\cdot\mathbf{S} + \mathbf{E}\cdot\mathbf{j}\right)$$

Since this equation is true for any volume (provided the particles remain inside it) we have

$$\frac{\partial u}{\partial t} + \nabla\cdot\mathbf{S} + \mathbf{E}\cdot\mathbf{j} = 0 \tag{10.4}$$

which gives the continuity equation for energy. It remains to find explicit expressions for u and $\mathbf{S}$. This can be done by relating the current to Maxwell's equation

$$-\frac{1}{c}\frac{\partial\mathbf{E}}{\partial t} + \nabla\times\mathbf{B} = \frac{4\pi}{c}\,\mathbf{j}$$

and therefore

$$\mathbf{j} = \frac{c}{4\pi}\left(\nabla\times\mathbf{B} - \frac{1}{c}\frac{\partial\mathbf{E}}{\partial t}\right) \tag{10.5}$$

If we insert this expression in Eq.(10.1) we find

$$\frac{dE_{\mathrm{p}}}{dt} = \frac{c}{4\pi}\int_V d\mathbf{x}\,\mathbf{E}\cdot\left(\nabla\times\mathbf{B} - \frac{1}{c}\frac{\partial\mathbf{E}}{\partial t}\right) \tag{10.6}$$

We further use the identity

$$\nabla\cdot(\mathbf{E}\times\mathbf{B}) = \mathbf{B}\cdot\nabla\times\mathbf{E} - \mathbf{E}\cdot\nabla\times\mathbf{B} = -\frac{1}{c}\mathbf{B}\cdot\frac{\partial\mathbf{B}}{\partial t} - \mathbf{E}\cdot\nabla\times\mathbf{B}$$

where in the last step we used the other Maxwell equation $\nabla\times\mathbf{E} = (-1/c)\partial\mathbf{B}/\partial t$. If we use this expression in Eq.(10.6) we can write

$$\frac{dE_{\mathrm{p}}}{dt} = -\frac{c}{4\pi}\int_V d\mathbf{x}\left\{\frac{1}{2c}\frac{\partial}{\partial t}\left(\mathbf{E}^2+\mathbf{B}^2\right) + \nabla\cdot(\mathbf{E}\times\mathbf{B})\right\} = -\int_V d\mathbf{x}\left\{\frac{\partial u}{\partial t} + \nabla\cdot\mathbf{S}\right\} \tag{10.7}$$

where we defined

$$u = \frac{1}{8\pi}(\mathbf{E}^2+\mathbf{B}^2) \tag{10.8}$$

$$\mathbf{S} = \frac{c}{4\pi}\,\mathbf{E}\times\mathbf{B} \tag{10.9}$$

The vector $\mathbf{S}$ is called the Poynting vector and describes the energy flux through a given surface. We now see that the continuity equation (10.4) now attains the explicit form

$$\frac{1}{8\pi}\frac{\partial}{\partial t}(\mathbf{E}^2+\mathbf{B}^2) + \frac{c}{4\pi}\,\nabla\cdot(\mathbf{E}\times\mathbf{B}) + \mathbf{j}\cdot\mathbf{E} = 0 \tag{10.10}$$

Our next task is to find similar expressions that describe the conservation of momentum. Since the momentum is conserved for each of the three spatial directions we expect three equations of the type (10.10). It is then not difficult to guess that these equations, together with the energy conservation law (10.10), will represent the divergence of a rank two tensor, which will be denoted as the energy-momentum tensor. Let us start again with the Lorentz force equation for the momentum of Eq.(6.41) which we repeat here for convenience a bit more explicitly

$$\frac{d\mathbf{p}}{dt} = q(\mathbf{E}(\mathbf{z}(t),t) + \frac{1}{c}\,\mathbf{v}\times\mathbf{B}(\mathbf{z}(t),t))$$

in which we indicated in the arguments that the fields are evaluated at the position $\mathbf{z}(t)$ of the particle. For N particles with charges q_j at positions $\mathbf{z}_j(t)$ we have

$$\frac{d\mathbf{P}_{\rm p}}{dt} = \sum_{j=1}^{N}\frac{d\mathbf{p}_j}{dt} = \sum_{j=1}^{N} q_j\mathbf{E}(\mathbf{z}_j(t),t) + \sum_{j=1}^{N}\frac{1}{c}\,\mathbf{v}\times\mathbf{B}(\mathbf{z}_j(t),t))$$

where we defined $\mathbf{P}_{\rm p} = \mathbf{p}_1 + \ldots + \mathbf{p}_N$ to be the total momentum of the particles. If we define the charge density of N particles as

$$\rho(\mathbf{x},t) = \sum_{j=1}^{N} q_j\,\delta(\mathbf{x}-\mathbf{z}_j(t))$$

and use the expression for the current of Eq.(10.2) we can write that the change in total momentum of N particles in a volume V is given by

$$\frac{d\mathbf{P}_{\rm p}}{dt} = \int_V d\mathbf{x}\,\left(\rho(\mathbf{x},t)\,\mathbf{E}(\mathbf{x},t) + \frac{1}{c}\,\mathbf{j}(\mathbf{x},t)\times\mathbf{B}(\mathbf{x},t)\right) \tag{10.11}$$

If we now again use the expression for the current (10.5) together the Maxwell equation

$$\rho(\mathbf{x},t) = \frac{1}{4\pi}\nabla\cdot\mathbf{E}$$

then we can rewrite Eq.(10.11) as

$$\frac{d\mathbf{P}_{\rm p}}{dt} = \frac{1}{4\pi}\int_V d\mathbf{x}\,(\nabla\cdot\mathbf{E})\,\mathbf{E} + \frac{1}{4\pi}\int_V d\mathbf{x}\,\left(\nabla\times\mathbf{B} - \frac{1}{c}\frac{\partial\mathbf{E}}{\partial t}\right)\times\mathbf{B} \tag{10.12}$$

We now use that

$$\frac{1}{c}\frac{\partial}{\partial t}(\mathbf{E}\times\mathbf{B}) = \frac{1}{c}\frac{\partial\mathbf{E}}{\partial t}\times\mathbf{B} + \frac{1}{c}\mathbf{E}\times\frac{\partial\mathbf{B}}{\partial t} = \frac{1}{c}\frac{\partial\mathbf{E}}{\partial t}\times\mathbf{B} - \mathbf{E}\times(\nabla\times\mathbf{E})$$

to rewrite Eq.(10.12) as

$$\begin{aligned}\frac{d\mathbf{P}_{\rm p}}{dt} &= \frac{1}{4\pi}\int_V d\mathbf{x}\,(\nabla\cdot\mathbf{E})\,\mathbf{E} - \frac{1}{4\pi c}\int_V d\mathbf{x}\,\frac{\partial}{\partial t}(\mathbf{E}\times\mathbf{B})\\ &\quad - \frac{1}{4\pi}\int_V d\mathbf{x}\,\left[\mathbf{E}\times(\nabla\times\mathbf{E}) + \mathbf{B}\times(\nabla\times\mathbf{B})\right]\end{aligned}$$

Using $\nabla\cdot\mathbf{B} = 0$ this can be written more symmetrically as

$$\begin{aligned}\frac{d\mathbf{P}_{\rm p}}{dt} &= -\frac{1}{4\pi c}\int_V d\mathbf{x}\,\frac{\partial}{\partial t}(\mathbf{E}\times\mathbf{B})\\ &\quad + \frac{1}{4\pi}\int_V d\mathbf{x}\,\left[\mathbf{E}(\nabla\cdot\mathbf{E}) - \mathbf{E}\times(\nabla\times\mathbf{E}) + \mathbf{B}(\nabla\cdot\mathbf{B}) - \mathbf{B}\times(\nabla\times\mathbf{B})\right]\end{aligned}$$

Now the last term can be written as a divergence. We have

$$(\mathbf{E}\times(\nabla\times\mathbf{E}))_i = \sum_{j,k}\epsilon_{ijk}E_j(\nabla\times\mathbf{E})_k = \sum_{j,k,l,m}\epsilon_{ijk}\epsilon_{klm}E_j\partial_l E_m$$
$$= \sum_{j,k,l,m}(\delta_{il}\delta_{jm}-\delta_{im}\delta_{jl})E_j\partial_l E_m = \sum_j E_j\partial_i E_j - E_j\partial_j E_i$$

and therefore

$$E_i(\nabla\cdot\mathbf{E}) - (\mathbf{E}\times(\nabla\times\mathbf{E}))_i = \sum_j E_i\partial_j E_j + \partial_i E_j - E_j\partial_j E_i = \sum_j \partial_j(E_iE_j - \frac{1}{2}\delta_{ij}\mathbf{E}^2)$$

Our equation for the momentum change therefore becomes

$$\frac{d\mathbf{P}_\mathrm{p}}{dt} = -\frac{1}{4\pi c}\int_V d\mathbf{x}\,\frac{\partial}{\partial t}(\mathbf{E}\times\mathbf{B})$$
$$+\frac{1}{4\pi}\sum_{j=1}^{3}\int_V d\mathbf{x}\,\frac{\partial}{\partial x^j}\left\{\frac{1}{2}\delta_{ij}(\mathbf{E}^2+\mathbf{B}^2) - (E_iE_j + B_iB_j)\right\} \quad (10.13)$$

From Eq.(10.11) and the fact the this equation is valid for any volume V we find the expression

$$(\rho\mathbf{E} + \frac{1}{c}\mathbf{j}\times\mathbf{B}) = -\frac{1}{4\pi}\frac{\partial}{\partial t}(\mathbf{E}\times\mathbf{B})_i$$
$$-\frac{1}{4\pi}\sum_{j=1}^{3}\frac{\partial}{\partial x^j}\left\{\frac{1}{2}\delta_{ij}(\mathbf{E}^2+\mathbf{B}^2) - (E_iE_j + B_iB_j)\right\} \quad (10.14)$$

The left hand side of this equation describes the change in momentum of the particle. Momentum conservation then implies that the right hand side describes an opposite change in momentum of the electromagnetic field. From the first term on the right hand side we see that the vector

$$\mathbf{u} = \frac{1}{4\pi c}\mathbf{E}\times\mathbf{B} \quad (10.15)$$

represents the momentum density in volume V whereas the last term in Eq.(10.14) represents the divergence of a momentum flow in the three separate directions. If we compare Eq.(10.15) to Eq. (10.9) we see that

$$\mathbf{u} = \frac{1}{c^2}\mathbf{S} \quad (10.16)$$

So the Poynting vector yields both the energy flow and (divided by c^2) the momentum density. It is now clear that we could combine the lefthand sides of Eqs.(10.7) and (10.13) into a momentum four-vector.

$$P_\mathrm{p} = (\frac{E_\mathrm{p}}{c}, \mathbf{P})$$

such that we can write Eq.(10.7) and (10.13) as

$$\frac{dP^\mu_\mathrm{p}}{dt} = -\int_V d\mathbf{x}\sum_{\nu=0}^{3}\frac{\partial T^{\mu\nu}_\mathrm{f}}{\partial x^\nu} \quad (10.17)$$

where we defined the energy momentum tensor $T_{\mathrm{f}}^{\mu\nu}$ of the field as

$$T_{\mathrm{f}}^{\mu\nu} = \frac{1}{4\pi}\times \tag{10.18}$$

$$\begin{pmatrix} \frac{1}{2}(\mathbf{E}^2+\mathbf{B}^2) & (\mathbf{E}\times\mathbf{B})_1 & (\mathbf{E}\times\mathbf{B})_2 & (\mathbf{E}\times\mathbf{B})_3 \\ (\mathbf{E}\times\mathbf{B})_1 & \frac{1}{2}(\mathbf{E}^2+\mathbf{B}^2)-E_1^2-B_1^2 & -E_1E_2-B_1B_2 & -E_2E_3-B_2B_3 \\ (\mathbf{E}\times\mathbf{B})_2 & -E_1E_2-B_1B_2 & \frac{1}{2}(\mathbf{E}^2+\mathbf{B}^2)-E_2^2-B_2^2 & -E_1E_3-B_1B_3 \\ (\mathbf{E}\times\mathbf{B})_3 & -E_2E_3-B_2B_3 & -E_1E_3-B_1B_3 & \frac{1}{2}(\mathbf{E}^2+\mathbf{B}^2)-E_3^2-B_3^2 \end{pmatrix}$$

We write the indices in $T_{\mathrm{f}}^{\mu\nu}$ as a superscript since we want to get a four-vector with an upper index after taking a divergence. Of course, it is not clear that by writing out the components that we are actually dealing with a tensor. It is not difficult to see that $T_{\mathrm{f}}^{\mu\nu}$ must be closely related to the electromagnetic field tensor. There are not so many options to relate the two. Since the fields $\mathbf{E}$ and $\mathbf{B}$ occur quadratically in the energy-momentum tensor one may guess that $T_{\mathrm{f}}^{\mu\nu}$ should be written as a product of electromagnetic field tensors. Then we have essentially one option

$$T_{\mathrm{f}}^{\mu\nu} = \alpha \sum_{\lambda=0}^{3} F^{\mu}_{\ \lambda} F^{\lambda\nu} + \beta\, g^{\mu\nu} \sum_{\lambda,\tau=0}^{3} F_{\lambda\tau}F^{\lambda\tau} \tag{10.19}$$

The last term in this equation was already evaluated in before in Eq.(6.47) and gives

$$\sum_{\lambda,\tau=0}^{3} F_{\lambda\tau}F^{\lambda\tau} = 2(\mathbf{E}^2-\mathbf{B}^2)$$

It therefore remains to evaluate the first term in Eq.(10.19). The explicit form of $F^{\lambda\nu}$ has been given directly after Eq.(6.47), whereas $F^{\mu}_{\ \lambda}$ is easily calculated from

$$F^{\mu}_{\ \lambda} = \sum_{\nu=0}^{3} g^{\mu\nu}F_{\nu\lambda} = g^{\mu\mu}F_{\mu\lambda}$$

since Minkowski metric is diagonal. Therefore

$$F^{\mu}_{\ \lambda} = \begin{pmatrix} 0 & E_1 & E_2 & E_3 \\ E_1 & 0 & B_3 & -B_2 \\ E_2 & -B_3 & 0 & B_1 \\ E_3 & B_2 & -B_1 & 0 \end{pmatrix}$$

and therefore

$$\begin{aligned} \sum_{\lambda=0}^{3} F^{\mu}_{\ \lambda}F^{\lambda\nu} &= \begin{pmatrix} 0 & E_1 & E_2 & E_3 \\ E_1 & 0 & B_3 & -B_2 \\ E_2 & -B_3 & 0 & B_1 \\ E_3 & B_2 & -B_1 & 0 \end{pmatrix} \begin{pmatrix} 0 & E_1 & E_2 & E_3 \\ -E_1 & 0 & B_3 & -B_2 \\ -E_2 & -B_3 & 0 & B_1 \\ -E_3 & B_2 & -B_1 & 0 \end{pmatrix} \\ &= \begin{pmatrix} -\mathbf{E}^2 & -E_2B_3+E_3B_2 & E_1B_3-E_3B_1 & -E_1B_2+E_2B_1 \\ -E_2B_3+E_3B_2 & E_1^2-B_2^2-B_3^3 & E_1E_2+B_1B_2 & E_1E_3+B_1B_3 \\ E_1B_3-E_3B_1 & E_1E_2+B_1B_2 & E_2^2-B_1^2-B_3^3 & E_2E_3+B_2B_3 \\ -E_1B_2+E_2B_1 & E_1E_3+B_1B_3 & E_2E_3+B_2B_3 & E_3^2-B_1^2-B_2^3 \end{pmatrix} \end{aligned}$$

We see that we are getting very close to Eq.(10.18). All off-diagonal elements are the same apart from the pre factor $-1/4\pi$. Let us therefore choose $\alpha = -1/4\pi$. Let us then look at the first diagonal element. Using Eq.(10.19) we have

$$T_{\mathrm{f}}^{00} = \frac{1}{8\pi}(\mathbf{E}^2+\mathbf{B}^2) = \frac{1}{4\pi}\mathbf{E}^2 + \beta\, g^{00}\, 2(\mathbf{B}^2-\mathbf{E}^2)$$

Since $g^{00} = -1$ we see that we must have $\beta = -1/16\pi$. You can check yourself that this also works out correctly for the other diagonal elements. We therefore find that

$$T_{\mathrm{f}}^{\mu\nu} = -\frac{1}{4\pi} \sum_{\lambda=0}^{3} F^{\mu}{}_{\lambda} F^{\lambda\nu} - \frac{1}{16\pi} g^{\mu\nu} \sum_{\lambda,\tau=0}^{3} F_{\lambda\tau} F^{\lambda\tau} \tag{10.20}$$

Now since F and g are tensors it follows that T_{f} is a tensor as well. It is called the *energy-momentum tensor* of the electromagnetic field.

10.2 The energy-momentum tensor in general coordinates

A nice feature of Eq.(10.20) is that it is valid in general space-times. In particular, we can derive in general coordinates that

$$\sum_{\nu=0}^{3} T_{\mathrm{f};\nu}^{\mu\nu} = -\frac{1}{c} \sum_{\nu=0}^{3} F^{\mu}{}_{\nu}\, j^{\nu} \tag{10.21}$$

which is equivalent to Eqs. (10.1) and (10.11). To prove this equation directly from Eq.(10.20) we need the following identity

$$F_{\mu\nu;\tau} + F_{\nu\tau;\mu} + F_{\tau\mu;\nu} = 0 \tag{10.22}$$

which is identical to the condition $dF = 0$. This is readily derived. If

$$F = \frac{1}{2} \sum_{\mu,\nu=0}^{3} F_{\mu\nu} dx^{\mu} \wedge dx^{\nu} = \sum_{\mu<\nu=0}^{3} F_{\mu\nu} dx^{\mu} \wedge dx^{\nu}$$

then

$$\begin{aligned} 0 = dF &= \sum_{\tau,\mu<\nu=0}^{3} \frac{\partial F_{\mu\nu}}{\partial x^{\tau}} dx^{\tau} \wedge dx^{\mu} \wedge dx^{\nu} = \sum_{\tau,\mu<\nu=0}^{3} F_{\mu\nu;\tau}\, dx^{\tau} \wedge dx^{\mu} \wedge dx^{\nu} \\ &= \sum_{\tau<\mu<\nu=0}^{3} F_{\mu\nu;\tau} \left[dx^{\tau} \wedge dx^{\mu} \wedge dx^{\nu} - dx^{\mu} \wedge dx^{\tau} \wedge dx^{\nu} + dx^{\mu} \wedge dx^{\nu} \wedge dx^{\tau} \right] \\ &= \sum_{\tau<\mu<\nu=0}^{3} \left[F_{\mu\nu;\tau} + F_{\tau\mu;\nu} + F_{\nu\tau;\mu} \right] dx^{\tau} \wedge dx^{\mu} \wedge dx^{\nu} \end{aligned}$$

which yields Eq.(10.22). Let us now calculate the divergence of the energy-momentum tensor. We have

$$\begin{aligned} \sum_{\nu} T_{\mathrm{f}\ ;\nu}^{\mu\nu} &= -\frac{1}{4\pi} \sum_{\lambda,\nu} F^{\mu}{}_{\lambda;\nu} F^{\lambda\nu} + F^{\mu}{}_{\lambda} F^{\lambda\nu}{}_{;\nu} - \frac{1}{16\pi} \sum_{\lambda,\tau,\nu} g^{\mu\nu} \left(F_{\lambda\tau;\nu} F^{\lambda\tau} + F_{\lambda\tau} F^{\lambda\tau}{}_{;\nu} \right) \\ &= -\frac{1}{4\pi} \sum_{\lambda,\alpha,\nu} g^{\mu\alpha} F_{\alpha\lambda;\nu} F^{\lambda\nu} - \frac{1}{4\pi} \sum_{\lambda} F^{\mu}{}_{\lambda} \frac{4\pi}{c} j^{\lambda} - \frac{1}{8\pi} \sum_{\lambda,\tau,\nu} g^{\mu\nu} F_{\lambda\tau;\nu} F^{\lambda\tau} \end{aligned} \tag{10.23}$$

where in the second term we used the Maxwell equations. We will now show that the first and last term cancel. We will manipulate the first term a bit. We have using Eq.(10.22)

$$\begin{aligned} -\frac{1}{4\pi} \sum_{\lambda,\alpha,\nu} g^{\mu\alpha} F_{\alpha\lambda;\nu} F^{\lambda\nu} &= -\frac{1}{8\pi} \sum_{\lambda,\alpha,\nu} g^{\mu\alpha} \left(F_{\alpha\lambda;\nu} F^{\lambda\nu} + F_{\alpha\nu;\lambda} F^{\nu\lambda} \right) \\ &= -\frac{1}{8\pi} \sum_{\lambda,\alpha,\nu} g^{\mu\alpha} \left(F_{\alpha\lambda;\nu} + F_{\nu\alpha;\lambda} \right) F^{\lambda\nu} = -\frac{1}{8\pi} \sum_{\lambda,\alpha,\nu} g^{\mu\alpha} \left(-F_{\lambda\nu;\alpha} F^{\lambda\nu} \right) \end{aligned}$$

which exactly cancels the last term in Eq.(10.23). We have therefore shown that the divergence of the energy-momentum tensor exactly produces minus the Lorentz force. The latter gives exactly the change in momentum of the charged particles. We may therefore wonder whether there would not be an energy-momentum tensor $T_{\mathrm{m}}^{\mu\nu}$ for matter as well, such that

$$\sum_{\nu} T_{\mathrm{f}\ ;\nu}^{\mu\nu} = -\sum_{\nu} T_{\mathrm{m}\ ;\nu}^{\mu\nu} \tag{10.24}$$

or, equivalently, if we define the total energy-momentum tensor for matter and field

$$T^{\mu\nu} = T_{\mathrm{f}}^{\mu\nu} + T_{\mathrm{m}}^{\mu\nu}$$

then

$$\sum_{\nu} T_{\ ;\nu}^{\mu\nu} = 0$$

which describes the conservation of energy and momentum of the matter and field system. This is indeed the case since even when the particles were neutral the four-momentum would be conserved and we therefore expect this four-momentum to be the divergence of a tensor. This should be compared to charge conservation. Since charge is a scalar quantity its conservation law is described as the divergence of a vector instead. Given the form of the current in Eq.(9.16) it is not so difficult to guess what $T_{\mathrm{m}}^{\mu\nu}$ could be. The simplest generalization of Eq.(9.16) for a single particle will be

$$T_{\mathrm{m}}^{\mu\nu} = mc \int d\tau \, \frac{dz^\mu}{d\tau} \frac{dz^\nu}{d\tau} \, \delta^{(4)}(x - z(\tau)) \tag{10.25}$$

whereas if we have several particles then

$$T_{\mathrm{m}}^{\mu\nu} = \sum_{j=1}^{N} m_j c \int d\tau \, \frac{dz_j^\mu}{d\tau} \frac{dz_j^\nu}{d\tau} \, \delta^{(4)}(x - z_j(\tau))$$

Let us see if Eq.(10.25) gives us what we expect. We can write with $x^0 = ct(\tau)$ that

$$\begin{aligned} T_{\mathrm{m}}^{\mu\nu} &= mc \int d\tau' \, \frac{dz^\mu}{d\tau'} \frac{dz^\nu}{d\tau'} \, \delta^{(3)}(\mathbf{x} - \mathbf{z}(\tau'))\delta(x^0 - z^0(\tau')) \\ &= mc \int d\tau' \, \frac{dz^\mu}{d\tau'} \frac{dz^\nu}{d\tau'} \, \delta^{(3)}(\mathbf{x} - \mathbf{z}(\tau'))\delta(\tau - \tau') \frac{1}{|\frac{dz^0}{d\tau'}|} \\ &= m \frac{dz^\mu}{dt} \frac{dt}{d\tau} \frac{dz^\nu}{dt} \delta^{(3)}(\mathbf{x} - \mathbf{z}(t)) = \gamma m \frac{dz^\mu}{dt} \frac{dz^\nu}{dt} \delta^{(3)}(\mathbf{x} - \mathbf{z}(t)) \end{aligned}$$

This means, for instance, that

$$\begin{aligned} T_{\mathrm{m}}^{\mu 0} &= \gamma m \frac{dz^\mu}{dt} \frac{dz^0}{dt} \, \delta^{(3)}(\mathbf{x} - \mathbf{z}(t)) = \gamma mc \frac{dz^\mu}{dt} \, \delta^{(3)}(\mathbf{x} - \mathbf{z}(t)) \\ &= (\gamma mc^2, \gamma mc \frac{d\mathbf{z}}{dt}) \, \delta^{(3)}(\mathbf{x} - \mathbf{z}(t)) \end{aligned}$$

So the first column of the energy-momentum tensor gives indeed the energy and momentum density as expected, and we have

$$\int d\mathbf{x} \, T_{\mathrm{m}}^{\mu 0}(\mathbf{x}, t) = \gamma mc \frac{dz^\mu}{dt} = (\gamma mc^2, \gamma mc \frac{d\mathbf{z}}{dt})$$

let us now check the divergence. We have

$$\begin{aligned}\sum_{\nu=0}^{3}\frac{\partial T_{\mathrm{m}}^{\mu\nu}}{\partial x^\nu} &= mc\int d\tau\,\frac{dz^\mu}{d\tau}\frac{dz^\nu}{d\tau}\,\frac{\partial}{\partial x^\nu}\delta^{(4)}(x-z(\tau))\\ &= -mc\int d\tau\,\frac{dz^\mu}{d\tau}\frac{d}{d\tau}\delta^{(4)}(x-z(\tau)) = mc\int d\tau\frac{d^2z^\mu}{d\tau^2}\,\delta^{(4)}(x-z(\tau))\\ &= c\int d\tau\,\frac{dp^\mu}{d\tau}\,\delta^{(4)}(x-z(\tau)) = \frac{dp^\mu}{dt}\delta^{(3)}(\mathbf{x}-\mathbf{z}(t))\end{aligned} \tag{10.26}$$

where p^μ is the four-momentum of the particle. For N particles it is clear that

$$\sum_{\nu=0}^{3}\frac{\partial T_{\mathrm{m}}^{\mu\nu}}{\partial x^\nu} = \sum_{j=1}^{N}\frac{dp_j^\mu}{dt}\delta^{(3)}(\mathbf{x}-\mathbf{z}_j(t))$$

The total momentum change P_{p} of the particles is then given by

$$\frac{dP_{\mathrm{p}}}{dt} = \sum_{\nu=0}^{3}\int_V d\mathbf{x}\,\frac{\partial T_{\mathrm{m}}^{\mu\nu}}{\partial x^\nu} = \sum_{j=1}^{N}\frac{dp_j^\mu}{dt}$$

Inserting this expression into the left hand side of Eq.(10.17) then gives

$$0 = \sum_{\nu=0}^{3}\frac{\partial}{\partial x^\nu}\left(T_{\mathrm{f}}^{\mu\nu}+T_{\mathrm{m}}^{\mu\nu}\right)$$

as expected. So far our discussion of the energy momentum tensor for the particles was entirely Minkowskian. It turns out that to use general coordinates we only slightly have to modify Eq.(10.25) to

$$T_{\mathrm{m}}^{\mu\nu} = \frac{mc}{\sqrt{|g|}}\int d\tau\,\frac{dz^\mu}{d\tau}\frac{dz^\nu}{d\tau}\,\delta^{(4)}(x-z(\tau)) \tag{10.27}$$

where g is the determinant of the metric tensor as before. In the case of Minkowskian coordinates $|g|=1$ and we get back Eq.(10.25). This is not so surprising since $\sqrt{|g|}$ is just the Jacobian of a coordinate transformation and the delta function transforms under coordinate transformations as

$$\delta(\mathbf{x}-\mathbf{x}') = \frac{1}{|\frac{\partial\mathbf{x}}{\partial\mathbf{y}}|}\delta(\mathbf{y}-\mathbf{y}')$$

where $|\partial\mathbf{x}/\partial\mathbf{y}| = \sqrt{|g|}$ is the Jacobian. For instance, we have in spherical coordinates

$$\delta(x-x')\delta(y-y')\delta(z-z') = \frac{1}{r^2\sin\theta}\delta(r-r')\,\delta(\theta-\theta')\delta(\phi-\phi')$$

Similarly, the general expression for the current of Eq.(9.16) becomes

$$j^\mu(x) = \frac{c\,q}{\sqrt{|g|}}\int d\tau\,\frac{dz^\mu}{d\tau}\,\delta^{(4)}(x-z(\tau)) \tag{10.28}$$

in general coordinates. We can check, for instance, that this current obeys

$$\begin{aligned}\sum_{\mu=0}^{3} j^\mu_{;\mu} &= \frac{1}{\sqrt{|g|}}\frac{\partial}{\partial x^\mu}(\sqrt{|g|}\,j^\mu) = \frac{cq}{\sqrt{|g|}}\int d\tau\,\frac{dz^\mu}{d\tau}\,\frac{\partial}{\partial x^\mu}\delta^{(4)}(x-z(\tau))\\ &= -\frac{cq}{\sqrt{|g|}}\int d\tau\,\frac{d}{d\tau}\delta^{(4)}(x-z(\tau)) = 0\end{aligned}$$

which is the charge conservation law. Let us now calculate the covariant divergence of $T_{\rm m}^{\mu\nu}$. For this it will be useful to use the formula

$$\begin{aligned}
\sum_\nu A^{\mu\nu}_{;\nu} &= \sum_\nu \frac{\partial A^{\mu\nu}}{\partial x^\nu} + \sum_{\nu\eta} A^{\eta\nu}\Gamma^\mu_{\nu\eta} + \sum_{\nu\eta} A^{\nu\eta}\Gamma^\nu_{\nu\eta} \\
&= \sum_\nu \frac{\partial A^{\mu\nu}}{\partial x^\nu} + \sum_{\nu\eta} A^{\eta\nu}\Gamma^\mu_{\nu\eta} + \sum_{\eta} A^{\mu\eta}\frac{1}{\sqrt{|g|}}\frac{\partial}{\partial x^\eta}\sqrt{|g|} \\
&= \frac{1}{\sqrt{|g|}}\sum_\nu \frac{\partial}{\partial x^\nu}(\sqrt{|g|}A^{\mu\nu}) + \sum_{\nu\eta} A^{\eta\nu}\Gamma^\mu_{\nu\eta}
\end{aligned}$$

which is just a modification of Eq.(7.34) for the case of a symmetric tensor. If we use this equation then we find that the energy-momentum tensor of the particles satisfies

$$\begin{aligned}
\sum_\nu \frac{\partial T_{\rm m}^{\mu\nu}}{\partial x^\nu} &= \frac{mc}{\sqrt{|g|}}\sum_\nu \int d\tau\,\frac{dz^\mu}{d\tau}\frac{dz^\nu}{d\tau}\,\frac{\partial}{\partial x^\nu}\delta^{(4)}(x-z(\tau)) + \frac{mc}{\sqrt{|g|}}\sum_{\nu\eta}\int d\tau\,\Gamma^\mu_{\nu\eta}\,\frac{dz^\nu}{d\tau}\frac{dz^\eta}{d\tau}\delta^{(4)}(x-z(\tau)) \\
&= -\frac{mc}{\sqrt{|g|}}\int d\tau\,\frac{dz^\mu}{d\tau}\frac{d}{d\tau}\delta^{(4)}(x-z(\tau)) + \frac{mc}{\sqrt{|g|}}\sum_{\nu\eta}\int d\tau\,\Gamma^\mu_{\nu\eta}\,\frac{dz^\nu}{d\tau}\frac{dz^\eta}{d\tau}\delta^{(4)}(x-z(\tau)) \\
&= \frac{mc}{\sqrt{|g|}}\int d\tau\left\{\frac{d^2z^\mu}{d\tau^2} + \sum_{\nu\eta}\Gamma^\mu_{\nu\eta}\,\frac{dz^\nu}{d\tau}\frac{dz^\eta}{d\tau}\right\}\delta^{(4)}(x-z(\tau)) \\
&= \frac{mc}{\sqrt{|g|}}\int d\tau\,\frac{Dp^\mu}{d\tau}\,\delta^{(4)}(x-z(\tau))
\end{aligned} \tag{10.29}$$

which is the generalization of Eq.(10.26) to general coordinates. If we now use the Lorentz force law

$$\frac{Dp^\mu}{d\tau} = \frac{q}{c}\sum_\nu F^\mu_{\ \nu}\frac{dz^\nu}{d\tau}$$

then we find that

$$\sum_\nu \frac{\partial T_{\rm m}^{\mu\nu}}{\partial x^\nu} = \frac{q}{\sqrt{|g|}}\sum_\nu \int d\tau\,F^\mu_{\ \nu}\frac{dz^\nu}{d\tau}\delta^{(4)}(x-z(\tau)) = \frac{1}{c}\sum_\nu F^\mu_{\ \nu}(x)j^\nu(x)$$

where we used the general identity

$$\int d\tau f(z(\tau))\,\delta^{(4)}(x-z(\tau)) = f(x)\int d\tau\,\delta^{(4)}(x-z(\tau))$$

and the definition of the current of Eq.(10.28). We therefore recovered minus the right hand side of Eq.(10.21). We find that if we define the energy-momentum tensor of N particles to be

$$T_{\rm m}^{\mu\nu} = \sum_{j=1}^N \frac{m_jc}{\sqrt{|g|}}\int d\tau\,\frac{dz_j^\mu}{d\tau}\frac{dz_j^\nu}{d\tau}\,\delta^{(4)}(x-z_j(\tau)) \tag{10.30}$$

then Eq.(10.24) is satisfied and we have

$$0 = \sum_\nu T^{\mu\nu}_{{\rm f}\ ;\nu} + \sum_\nu T^{\mu\nu}_{{\rm m}\ ;\nu} \tag{10.31}$$

which expresses the laws of energy and momentum conservation of the combined field-matter system in general coordinates. In fact, as is clear form our derivation, Eq.(10.31) implies the Lorentz force law if we specify Maxwell's equations. So we can either use the Maxwell equations and the Lorentz force law to derive Eq.(10.31) or assume Maxwell's equations and the Eq.(10.31) to derive the Lorentz force law.

Chapter 11

Radiation by moving charges

In this Chapter we derive the angular distribution of radiated power by accelerated charges. We further derive how the total radiated power depends on the velocity and acceleration leading to the famous Larmor formula. We further derive the spectral distribution of radiated power for a charge in circular motion.

11.1 Radiated power

Let us summarize again the formulas of Chapter 9 for the electric and magnetic fields produced by a charge q following a trajectory $\mathbf{z}(t')$. The electric and magnetic fields are given by

$$\mathbf{E} = \frac{q}{\kappa^3 R}\left\{\frac{1}{c}\,\mathbf{n}\times[(\mathbf{n}-\boldsymbol{\beta})\times\dot{\boldsymbol{\beta}}] + (1-|\boldsymbol{\beta}|^2)\frac{1}{R}(\mathbf{n}-\boldsymbol{\beta})\right\} \tag{11.1}$$

$$\mathbf{B} = \mathbf{n}\times\mathbf{E} \tag{11.2}$$

where

$$R = |\mathbf{x}-\mathbf{z}(t')| \quad , \quad \mathbf{n}(t') = \frac{\mathbf{x}-\mathbf{z}(t')}{|\mathbf{x}-\mathbf{z}(t')|}$$

$$\kappa = 1 - \frac{1}{c}\,\mathbf{n}(t')\cdot\dot{\mathbf{z}}(t')$$

$$\boldsymbol{\beta} = \frac{1}{c}\,\dot{\mathbf{z}}(t')$$

and t' is the retarded time determined by the equation

$$c(t-t') = |\mathbf{x}-\mathbf{z}(t')| \tag{11.3}$$

If the charge is accelerated it will loose energy. Let us see how we can calculate this energy loss. In Eq.(10.9) we saw that the Poynting vector $\mathbf{S}$ describes the energy flux. Hence $\mathbf{S}\cdot\mathbf{n}$ is the energy flux per unit area per unit time detected at a point $\mathbf{x}$ at time t of radiation emitted by the charge q at retarded time t'.

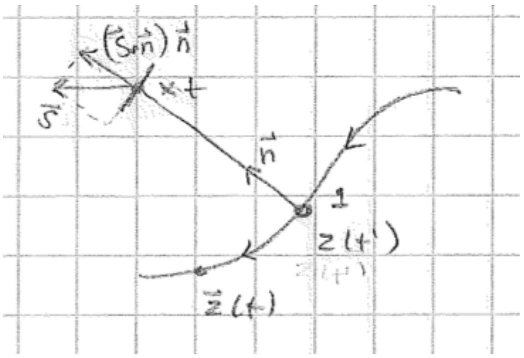

Figure 11.1: Poyting vector $\mathbf{S}$ projected on the direction $\mathbf{n}$.

The energy radiated during a period of acceleration from t'_1 to t'_2 is measured in $\mathbf{x}$ between times t_1 and t_2 where

$$t_i = t'_i + \frac{|\mathbf{x} - \mathbf{z}(t'_i)|}{c}$$

and $i = 1, 2$ is then given by

$$E = \int_{t_1}^{t_2} dt\, (\mathbf{S} \cdot \mathbf{n})(t) = \int_{t'_1}^{t'_2} dt'\, \frac{dt}{dt'}\, (\mathbf{S} \cdot \mathbf{n})(t') = \int_{t'_1}^{t'_2} dt'\, \kappa\, (\mathbf{S} \cdot \mathbf{n})(t') \tag{11.4}$$

where we with some abuse of notation wrote $(\mathbf{S} \cdot \mathbf{n})(t(t')) = (\mathbf{S} \cdot \mathbf{n})(t')$, i.e. we can regard the Poynting vector in $\mathbf{x}$ projected on $\mathbf{n}$ also as a function of the retarded time t'. We further used from Eq.(11.3) that for fixed $\mathbf{x}$

$$c(\frac{dt}{dt'} - 1) = -\frac{d\mathbf{z}}{dt'} \cdot \mathbf{n}(t') \implies \frac{dt}{dt'} = 1 - \frac{1}{c}\frac{d\mathbf{z}}{dt'} \cdot \mathbf{n}(t') = \kappa \tag{11.5}$$

We therefore find that the quantity $\kappa\, (\mathbf{S} \cdot \mathbf{n})(t')$ is the instantaneously radiated power, i.e. the rate of change of energy due to radiation that will pass through a unit area at position $\mathbf{x}$.

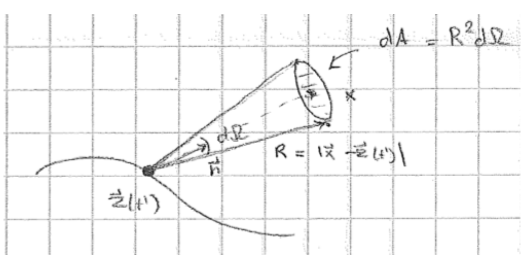

Figure 11.2: Radiated power in a solid angle $d\Omega$ around direction $\mathbf{n}$.

It will be useful to introduce another physical quantity. According to Eq.(11.4) the power $dP(\mathbf{n})$ radiated into a solid angle $d\Omega$ in the direction of $\mathbf{n}$ is given by

$$dP(\mathbf{n}) = \kappa\, (\mathbf{S} \cdot \mathbf{n})(t')\, R^2\, d\Omega$$

and we therefore define the function

$$\frac{dP}{d\Omega} = \kappa\, R^2\, (\mathbf{S} \cdot \mathbf{n})(t') \tag{11.6}$$

to be the radiated power emitted at time t' in direction $\mathbf{n}$. The Poynting vector has the explicit form

$$\mathbf{S} = \frac{c}{4\pi}\mathbf{E}\times\mathbf{B} \tag{11.7}$$

and from Eqs.(11.1) and (11.2) we there can therefore write

$$\mathbf{S} = \frac{c}{4\pi}\mathbf{E}\times(\mathbf{n}\times\mathbf{E}) = \frac{c}{4\pi}((\mathbf{E}\cdot\mathbf{E})\,\mathbf{n} - \mathbf{E}\,(\mathbf{E}\cdot\mathbf{n}))$$

From Eq.(11.6) we can therefore write

$$\frac{dP}{d\Omega} = \kappa\, R^2\,\left(\mathbf{E}^2 - (\mathbf{E}\cdot\mathbf{n})^2\right) \tag{11.8}$$

Now we can use Eq.(11.1) which we will rewrite as

$$\mathbf{E} = \frac{\mathbf{a}}{R} + \frac{\mathbf{b}}{R^2}$$

where

$$\begin{aligned} \mathbf{a} &= \frac{q}{\kappa^3\,c}\,\mathbf{n}\times((\mathbf{n}-\boldsymbol{\beta})\times\dot{\boldsymbol{\beta}}) \\ \mathbf{b} &= \frac{q}{\kappa^3}\,(1-|\boldsymbol{\beta}|^2)\,(\mathbf{n}-\boldsymbol{\beta}) \end{aligned}$$

Then

$$\begin{aligned} \mathbf{E}^2 &= \frac{\mathbf{a}^2}{R^2} + \frac{2\mathbf{a}\cdot\mathbf{b}}{R^3} + \frac{\mathbf{b}^2}{R^4} \\ \mathbf{E}\cdot\mathbf{n} &= \frac{\mathbf{b}\cdot\mathbf{n}}{R^2} \end{aligned}$$

where we used that $\mathbf{a}\cdot\mathbf{n} = 0$. We then have that

$$\mathbf{E}^2 - (\mathbf{E}\cdot\mathbf{n})^2 = \frac{\mathbf{a}^2}{R^2} + \frac{2\mathbf{a}\cdot\mathbf{b}}{R^3} + \frac{1}{R^4}(\mathbf{b}^2 - (\mathbf{b}\cdot\mathbf{n})^2)$$

When we insert this expression into Eq.(11.8) we see that the last two terms do not lead to a net energy flux through a surface at large distances. Therefore only the first term contributes to the radiation. Inserting this into Eq.(11.8) then gives

$$\frac{dP}{d\Omega} = \frac{c\kappa}{4\pi}\mathbf{a}^2 = \frac{q^2}{4\pi c}\frac{|\mathbf{n}\times((\mathbf{n}-\boldsymbol{\beta})\times\dot{\boldsymbol{\beta}})|^2}{\kappa^5} = \frac{q^2}{4\pi c}\frac{|\mathbf{n}\times((\mathbf{n}-\boldsymbol{\beta})\times\dot{\boldsymbol{\beta}})|^2}{(1-\mathbf{n}\cdot\boldsymbol{\beta})^5}$$

If we expand the outer products in the numerator we have

$$\begin{aligned} &|\mathbf{n}\times((\mathbf{n}-\boldsymbol{\beta})\times\dot{\boldsymbol{\beta}})|^2 = |(\mathbf{n}-\boldsymbol{\beta})(\mathbf{n}\cdot\dot{\boldsymbol{\beta}}) - \dot{\boldsymbol{\beta}}(1-\boldsymbol{\beta}\cdot\mathbf{n})|^2 \\ &= (\mathbf{n}\cdot\dot{\boldsymbol{\beta}})^2\,(1-2\,\mathbf{n}\cdot\boldsymbol{\beta}+\boldsymbol{\beta}^2) - 2(\mathbf{n}\cdot\dot{\boldsymbol{\beta}} - \boldsymbol{\beta}\cdot\dot{\boldsymbol{\beta}})(1-\boldsymbol{\beta}\cdot\mathbf{n})(\mathbf{n}\cdot\dot{\boldsymbol{\beta}}) + \dot{\boldsymbol{\beta}}^2(1-\boldsymbol{\beta}\cdot\mathbf{n})^2 \\ &= (\mathbf{n}\cdot\dot{\boldsymbol{\beta}})^2(\boldsymbol{\beta}^2-1) + 2(\mathbf{n}\cdot\dot{\boldsymbol{\beta}})(\boldsymbol{\beta}\cdot\dot{\boldsymbol{\beta}})(1-\boldsymbol{\beta}\cdot\mathbf{n}) + \dot{\boldsymbol{\beta}}^2(1-\boldsymbol{\beta}\cdot\mathbf{n})^2 \end{aligned}$$

and we can therefore write

$$\frac{dP}{d\Omega} = \frac{q^2}{4\pi c}\left\{\frac{\dot{\boldsymbol{\beta}}^2}{(1-\boldsymbol{\beta}\cdot\mathbf{n})^3} + \frac{2(\mathbf{n}\cdot\dot{\boldsymbol{\beta}})(\boldsymbol{\beta}\cdot\dot{\boldsymbol{\beta}})}{(1-\boldsymbol{\beta}\cdot\mathbf{n})^4} - (1-\boldsymbol{\beta}^2)\frac{(\mathbf{n}\cdot\dot{\boldsymbol{\beta}})^2}{(1-\boldsymbol{\beta}\cdot\mathbf{n})^5}\right\} \tag{11.9}$$

11.2 Angular distribution and Larmor's formula

From expression (11.9) we can calculate the angular distribution of the instantaneously radiated power.

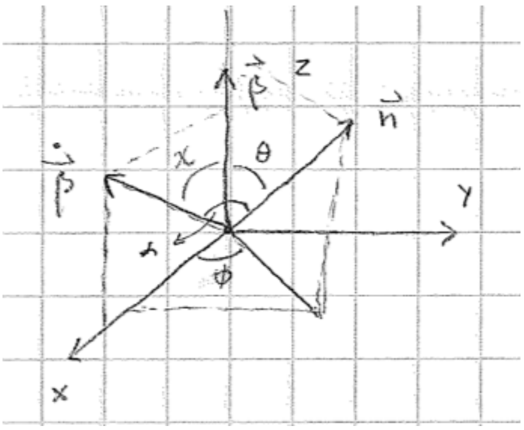

Figure 11.3: Angles between the vectors $\boldsymbol{\beta}$, $\dot{\boldsymbol{\beta}}$ and $\mathbf{n}$.

We can always choose a coordinate system where the vector $\boldsymbol{\beta}$ is pointing along the z-axis and the acceleration vector $\dot{\boldsymbol{\beta}}$ is lying in the $x-z$-plane. We observe radiation in a direction $\mathbf{n}$ given by

$$\mathbf{n} = \begin{pmatrix} \cos\phi\sin\theta \\ \sin\phi\sin\theta \\ \cos\theta \end{pmatrix}$$

The vector $\dot{\boldsymbol{\beta}}$ is given by

$$\dot{\boldsymbol{\beta}} = \dot{(\beta)} \begin{pmatrix} \sin\chi \\ 0 \\ \cos\chi \end{pmatrix} \tag{11.10}$$

where $\dot{\beta} = |\dot{\beta}|$. We see that

$$\mathbf{n}\cdot\boldsymbol{\beta} = \cos\theta \quad , \quad \boldsymbol{\beta}\cdot\dot{\boldsymbol{\beta}} = \beta\dot{\beta}\,\cos\chi$$

where $\beta = |\boldsymbol{\beta}|$. Further, if α is the angle between $\mathbf{n}$ and $\dot{\boldsymbol{\beta}}$ then we have

$$\mathbf{n}\cdot\dot{\boldsymbol{\beta}} = \dot{\beta}\,\cos\alpha = \dot{\beta}\,(\cos\phi\sin\theta\sin\chi + \cos\theta\cos\chi) \tag{11.11}$$

In terms of these angles we can now rewrite Eq.(11.9) as

$$\frac{dP}{d\Omega} = \frac{q^2\dot{\beta}^2}{4\pi c}\left\{\frac{1}{(1-\beta\,\cos\theta)^3} + \frac{2\beta\cos\chi\cos\alpha}{(1-\beta\,\cos\theta)^4} - (1-\beta^2)\frac{(\cos\alpha)^2}{(1-\beta\,\cos\theta)^5}\right\} \tag{11.12}$$

Let us consider some special cases. One obvious special case is the linear acceleration for which $\boldsymbol{\beta} \parallel \dot{\boldsymbol{\beta}}$, i.e. $\chi = 0, \pi$ and $\cos\alpha = \pm\cos\theta$. Then Eq.(11.12) gives

$$\frac{dP}{d\Omega} = \frac{q^2\dot{\beta}^2}{4\pi c}\frac{\sin^2\theta}{(1-\beta\,\cos\theta)^5} \tag{11.13}$$

The denominator never becomes zero since $\beta < 1$ but it can become small for $\beta \approx 1$. It is then clear that $dP/d\Omega$ is large for small angles θ and therefore the maximum of the power is radiated in the forward direction. We thus have

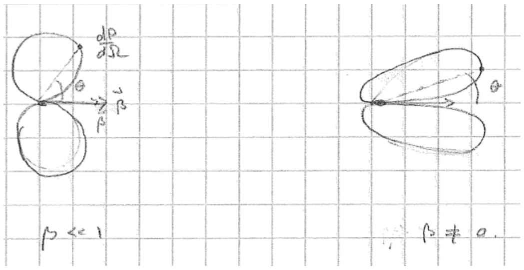

Figure 11.4: Angular distribution of radiation around a linearly accelerated charge.

in which the value of $dP/d\Omega$ for a given angle θ is plotted as the distance to the origin in a polar plot. The angle θ_m of maximum intensity is found by differentiating Eq.(11.13) with respect to $x = \cos\theta$. This gives

$$0 = \frac{d}{dx}\left(\frac{1-x^2}{(1-\beta x)^5}\right) \implies x = \frac{\pm\sqrt{1+15\beta^2}-1}{3\beta} \implies \theta_m = \arccos\frac{\sqrt{1+15\beta^2}-1}{3\beta}$$

(the other zero leads to $x < -1$). We have

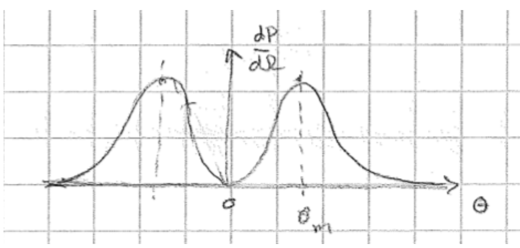

Figure 11.5: Angular distribution of radiation.

In the limit $\beta \to 1$ we have $\theta_m \to 0$. The other special case is when $\boldsymbol{\beta} \perp \dot{\boldsymbol{\beta}}$, i.e $\chi = \pi/2$. In that case we have from Eq.(11.11) that

$$\cos\alpha = \cos\phi\sin\theta \tag{11.14}$$

The case $\boldsymbol{\beta} \perp \dot{\boldsymbol{\beta}}$ represents physically the case of instantaneous circular motion

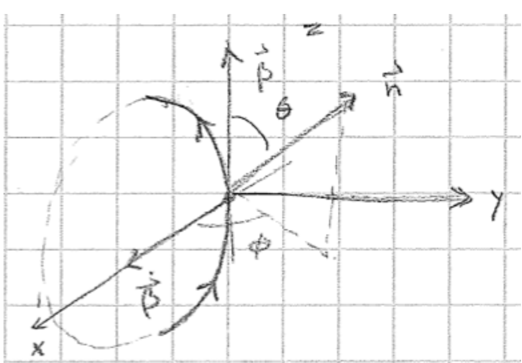

Figure 11.6: Instanteneous circular motion.

Inserting $\chi = \pi/2$ and Eq.(11.14) into Eq.(11.12) then yields

$$\frac{dP}{d\Omega} = \frac{q^2\dot{\beta}^2}{4\pi c}\left\{\frac{1}{(1-\beta\,\cos\theta)^3} - (1-\beta^2)\frac{\cos^2\phi\sin^2\theta}{(1-\beta\,\cos\theta)^5}\right\} \tag{11.15}$$

In this case the function $dP/d\Omega$ is maximal for $\theta = 0$. In the $x-z$-plane ($\phi = 0$) the function is zero whenever

$$1 = (1-\beta^2)\frac{\sin^2\theta}{(1-\beta\,\cos\theta)^2} \implies \cos\theta = \beta \implies \theta_0 = \arccos\beta$$

The angle $\theta_0 \to 0$ for $\beta \to 1$ so the radiated power is sharply peaked in the forward direction in this limit.

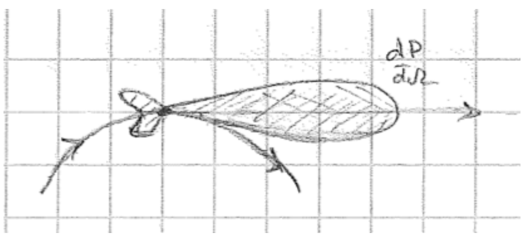

Figure 11.7: Radiation from circular motion.

Let us now calculate the total radiated power P in all directions. In that case we must integrate Eq.(11.12) over the angles ϕ and θ.

$$P = \int d\Omega\,\frac{dP}{d\Omega} = \int_0^\pi d\theta\int_0^{2\pi} d\phi\,\sin\theta\,\frac{dP}{d\Omega} \tag{11.16}$$

Let us first integrate over ϕ. Since only $\cos\alpha$ depends on ϕ we need to calculate

$$\int_0^{2\pi} d\phi\,\cos\alpha = \int_0^{2\pi} d\phi\,(\cos\phi\sin\theta\sin\chi + \cos\theta\cos\chi) = 2\pi\cos\theta\cos\chi$$
$$\int_0^{2\pi} d\phi\,\cos^2\alpha = \int_0^{2\pi} d\phi\,(\cos\phi\sin\theta\sin\chi + \cos\theta\cos\chi)^2 = \pi(\sin^2\theta\sin^2\chi + 2\cos^2\theta\cos^2\chi)$$

Using these expressions we have

$$P = \int_0^\pi d\theta\,\sin\theta\frac{q^2\dot{\beta}^2}{4\pi c}\left\{\frac{2\pi}{(1-\beta\cos\theta)^3} + \frac{4\pi\beta\cos^2\chi\cos\theta}{(1-\beta\cos\theta)^4}\right.$$
$$\left. + (\beta^2-1)\frac{\pi(\sin^2\theta\sin^2\chi + 2\cos^2\theta\cos^2\chi)}{(1-\beta\cos\theta)^5}\right\}$$

With the substitution $x = 1-\beta\cos\theta$ this integral can be carried out in a straightforward way. We find

$$P = \frac{2q^2\dot{\beta}^2}{3c}\frac{1-\beta^2\sin^2\chi}{(1-\beta^2)^3} \tag{11.17}$$

This formula can be written in a different form if we use

$$|\boldsymbol{\beta}\times\dot{\boldsymbol{\beta}}|^2 = \beta^2\dot{\beta}^2\,\sin^2\chi$$

and write

$$P = \frac{2q^2}{3c}\frac{1}{(1-\beta^2)^3}\left[\dot\beta^2 - |\boldsymbol\beta\times\dot{\boldsymbol\beta}|^2\right] \tag{11.18}$$

This is the relativistic generalization of the so-called *Larmor formula*. It is often given in the non-relativistic limit $\beta \ll 1$ when it attains the form

$$P = \frac{2q^2\dot\beta^2}{3c} = \frac{2q^2a^2}{3c^3}$$

where a is the acceleration of the particle. We can compare Eq.(11.17) for the case that the acceleration is parallel ($\chi = 0$) or perpendicular ($\chi = \pi/2$) to the velocity. We have

$$P_\parallel = P(\chi = 0) = \frac{2q^2\dot\beta^2}{3c}\frac{1}{(1-\beta^2)^3} \tag{11.19}$$

$$P_\perp = P(\chi = \pi/2) = \frac{2q^2\dot\beta^2}{3c}\frac{1}{(1-\beta^2)^2} \tag{11.20}$$

We see that, for a given acceleration $\dot\beta$ a factor $\gamma^2 = 1/(1-\beta^2)$ more radiation is produced when the acceleration is parallel to the velocity as compared to the case when it perpendicular. A different picture, however, arises when we consider the forces responsible for the acceleration. We have

$$\mathbf{F} = \frac{d\mathbf{p}}{dt} = \frac{d}{dt}m\gamma\mathbf{v} = m\gamma\frac{d\mathbf{v}}{dt} + m\mathbf{v}\gamma^3\frac{\mathbf{v}}{c^2}\cdot\frac{d\mathbf{v}}{dt}$$

where in the last term we differentiated $\gamma = (1 - v^2/c^2)^{-1/2}$. In the case that $\mathbf{v} \perp d\mathbf{v}/dt$ we see that

$$\mathbf{F} = m\gamma\frac{d\mathbf{v}}{dt}$$

However, if $\mathbf{v} \parallel d\mathbf{v}/dt$ then

$$\mathbf{F} = m\gamma\frac{d\mathbf{v}}{dt} + m\mathbf{v}\gamma^3\frac{v}{c^2}\frac{dv}{dt} = m\gamma(1 + \frac{v^2}{c^2}\gamma^2)\frac{d\mathbf{v}}{dt} = m\gamma^3\frac{d\mathbf{v}}{dt}$$

We therefore see that

$$\begin{aligned}\frac{d\mathbf{v}}{dt} &= \frac{1}{m\gamma}\frac{d\mathbf{p}}{dt} \qquad\qquad \mathbf{v}\perp\frac{d\mathbf{v}}{dt}\\ \frac{d\mathbf{v}}{dt} &= \frac{1}{m\gamma^3}\frac{d\mathbf{p}}{dt} \qquad\qquad \mathbf{v}\parallel\frac{d\mathbf{v}}{dt}\end{aligned}$$

Inserting these expression into Eqs.(11.19) and (11.20) then gives

$$\begin{aligned}P_\parallel &= \frac{2q^2\dot v^2}{3m^2c^3}\left(\frac{d\mathbf{p}}{dt}\right)^2\\ P_\perp &= \frac{2q^2\dot v^2}{3m^2c^3}\,\gamma^2\left(\frac{d\mathbf{p}}{dt}\right)^2\end{aligned}$$

We therefore see that, for a given applied force, a factor γ^2 more radiation is produced if the force is perpendicular to $\mathbf{v}$ as compared to the case in which it is parallel. In the former case the force is usually produced by a magnetic field while in the latter case it produced by an electric field.

11.3 Radiation spectrum for circular motion

We will now study at which frequencies and intensities a charged in circular motion will radiate. Let a particle with charge q carry out a circular motion in the $x-y$-plane, for instance due to the presence of a magnetic field.

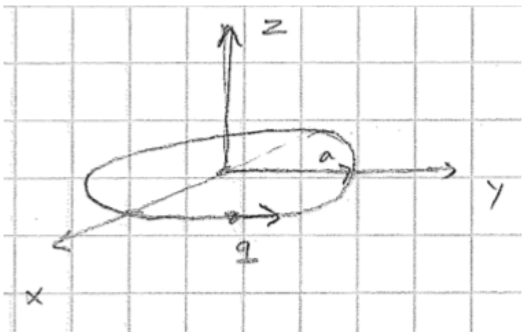

Figure 11.8: Charge moving along a circle in the plane $z=0$.

The path of the particle is given by

$$\mathbf{z}(t') = a \begin{pmatrix} \cos(\omega_0 t') \\ \sin(\omega_0 t') \\ 0 \end{pmatrix}$$

where a is the radius of the circle. For an observer at $\mathbf{x}$ the vector potential of the field generated by the charge is given by the Liénard-Wiechert potential

$$\mathbf{A}(\mathbf{x}, t) = \frac{q}{c} \frac{d\mathbf{z}(t')}{dt'} \frac{1}{\kappa \, |\mathbf{x} - \mathbf{z}(t')|} \tag{11.21}$$

The motion is periodic with period $T = 2\pi/\omega_0$

$$\mathbf{z}(t' + T) = \mathbf{z}(t')$$

and therefore the motion is periodic with the same period in the observation point $\mathbf{x}$ as can be checked directly as well from the equation for the retarded time. We therefore have

$$\mathbf{A}(\mathbf{x}, t) = \mathbf{A}(\mathbf{x}, t + T)$$

Because of this periodicity it will therefore be convenient to expand $\mathbf{A}$ in a Fourier series

$$\mathbf{A}(\mathbf{x}, t) = \sum_{l=-\infty}^{+\infty} \mathbf{A}_l(\mathbf{x}) \, e^{-i\omega_0 l t}$$

where the Fourier coefficient is given by

$$\mathbf{A}_l(\mathbf{x}) = \frac{1}{T} \int_0^T dt \, \mathbf{A}(\mathbf{x}, t) \, e^{i\omega_0 l t} \tag{11.22}$$

Since the vector potential $\mathbf{A}(\mathbf{x}, t)$ is real we see that $\mathbf{A}_l^*(\mathbf{x}) = \mathbf{A}_{-l}(\mathbf{x})$. We want to study the relative intensities of the various l-components which correspond to the higher harmonics

$l\omega_0$ of the fundamental frequency ω_0. We can now insert the explicit form of Eq.(11.21) into Eq.(11.22) to obtain

$$\mathbf{A}_l(\mathbf{x}) = \frac{q}{cT}\int_0^T dt\, \frac{d\mathbf{z}(t')}{dt'}\, \frac{1}{\kappa\, |\mathbf{x}-\mathbf{z}(t')|}\, e^{i\omega_0 l t} = \frac{q}{cT}\int_0^T dt'\, \frac{d\mathbf{z}(t')}{dt'}\, \frac{1}{|\mathbf{x}-\mathbf{z}(t')|}\, e^{i\omega_0 l t} \quad (11.23)$$

where we used that $dt'/dt = 1/\kappa$ (see Eq.(9.37)) and the fact that the period is also T in the t'-coordinate. To carry out this integral we need to know the dependence of t on t'. If we now use Eq.(11.5) we can write

$$\begin{aligned}\frac{dt}{dt'} &= 1 - \frac{1}{c}\frac{d\mathbf{z}}{dt'}\cdot \mathbf{n}(t') = 1 - \frac{a}{c}\begin{pmatrix} -\omega_0 \sin(\omega_0 t') \\ \omega_0\cos(\omega_0 t') \\ 0 \end{pmatrix}\cdot \begin{pmatrix} n_1 \\ n_2 \\ n_3 \end{pmatrix} \\ &= 1 - \frac{\omega_0 a}{c}\left[-\sin(\omega_0 t') n_1(t') + \cos(\omega_0 t') n_2(t')\right] \end{aligned} \quad (11.24)$$

This equation can be integrated since we know the explicit form of

$$\mathbf{n}(t') = \frac{\mathbf{x}-\mathbf{z}(t')}{|\mathbf{x}-\mathbf{z}(t')|}$$

(which would simply lead to the equation of the retarded time). However, if we are only interested in the radiation at large distance $|\mathbf{x}| \to \infty$ we do not need this relation in full generality. In that case the unit vector becomes constant and equal to $\mathbf{n}(t') \approx \mathbf{x}/|\mathbf{x}|$. Furthermore since the problem is rotationally symmetric under rotation around the z-axis no generality is lost by putting $\mathbf{x}$ in the $y-z$-plane and we can write

$$\mathbf{n} = \begin{pmatrix} 0 \\ \sin\theta \\ \cos\theta \end{pmatrix}$$

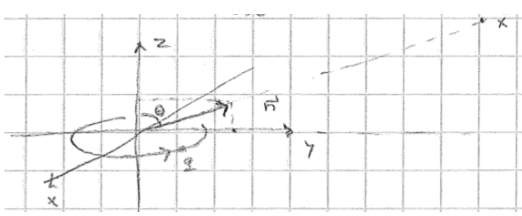

Figure 11.9: The observation point $\mathbf{x}$ is chosen in the plane $x = 0$.

Then Eq.(11.24) becomes

$$\frac{dt}{dt'} = 1 - \frac{\omega_0 a}{c}\cos(\omega_0 t')\sin\theta$$

This is now easily integrated to yield

$$t(t') = t' - \frac{a}{c}\sin\theta\,\sin(\omega_0 t') + \frac{R(0)}{c} \quad (11.25)$$

where we used that $t(0) = |\mathbf{x}-\mathbf{z}(0)|/c = R(0)/c$. We can now insert this into Eq.(11.23) to obtain

$$\mathbf{A}_l(\mathbf{x}) = \frac{q}{cT}e^{i\omega_0 l R(0)/c}\int_0^T dt'\, \frac{d\mathbf{z}(t')}{dt'}\, \frac{1}{|\mathbf{x}-\mathbf{z}(t')|}\, e^{i\omega_0 l (t' - \frac{a}{c}\sin\theta\, \sin(\omega_0 t'))}$$

Again since we take $|\mathbf{x}| \to \infty$ we have $|\mathbf{x} - \mathbf{z}(t')| \approx |\mathbf{x}|$ and $R(0) \approx |\mathbf{x}|$ we can write this as

$$\mathbf{A}_l(\mathbf{x}) = \frac{q}{cT} \frac{e^{ik|\mathbf{x}|}}{|\mathbf{x}|} \int_0^T dt' \, \frac{d\mathbf{z}(t')}{dt'} \, e^{i\omega_0 l (t' - \frac{a}{c} \sin\theta \, \sin(\omega_0 t'))}$$

where we used the abbreviation $k = \omega_0 l/c$. We now only need to insert the explicit form of $\mathbf{z}(t')$ and to perform the integral. This gives the following integrals for the x and y components of $\mathbf{A}_l$ (the z-component is zero)

$$\begin{aligned} A_{\mathrm{x}l} &= -\frac{q\omega_0 a}{cT} \frac{e^{ik|\mathbf{x}|}}{|\mathbf{x}|} \int_0^T dt' \, \sin(\omega_0 t') \, e^{i\omega_0 l (t' - \frac{a}{c} \sin\theta \, \sin(\omega_0 t'))} \\ A_{\mathrm{y}l} &= \frac{q\omega_0 a}{cT} \frac{e^{ik|\mathbf{x}|}}{|\mathbf{x}|} \int_0^T dt' \, \cos(\omega_0 t') \, e^{i\omega_0 l (t' - \frac{a}{c} \sin\theta \, \sin(\omega_0 t'))} \end{aligned}$$

Using the substitution $\varphi = \omega_0 t'$ we can write this as

$$A_{\mathrm{x}l} = -\frac{qa}{cT} \frac{e^{ik|\mathbf{x}|}}{|\mathbf{x}|} \int_0^{2\pi} d\varphi \, \sin\varphi \, e^{i(l\varphi - \eta l \sin\theta \, \sin\varphi)} \tag{11.26}$$

$$A_{\mathrm{y}l} = \frac{qa}{cT} \frac{e^{ik|\mathbf{x}|}}{|\mathbf{x}|} \int_0^{2\pi} d\varphi \, \cos\varphi \, e^{i(l\varphi - \eta l \sin\theta \, \sin\varphi)} \tag{11.27}$$

$$\tag{11.28}$$

where we defined $\eta = \omega_0 a/c$. These integrals are closely related to the Bessel functions defined by

$$J_l(x) = \frac{1}{2\pi} \int_0^{2\pi} d\varphi \, e^{i(l\varphi - x\sin\varphi)}$$

From this integral we see that

$$\int_0^{2\pi} d\varphi \, \sin\varphi \, e^{i(l\varphi - x\sin\varphi)} = 2\pi i \, \frac{d}{dx} J_l(x)$$

and further that

$$\begin{aligned} \int_0^{2\pi} d\varphi \, \cos\varphi \, e^{i(l\varphi - x\sin\varphi)} &= \frac{i}{x} \int_0^{2\pi} d\varphi \, e^{il\varphi} \frac{d}{d\varphi} e^{-ix\sin\varphi} = -\frac{i}{x} \int_0^{2\pi} d\varphi \, e^{-ix\sin\varphi} \frac{d}{d\varphi} e^{il\varphi} \\ &= \frac{2\pi l}{x} J_l(x) \end{aligned}$$

We therefore see that Eqs.(11.26) and (11.27) can be expressed in terms of the Bessel functions and their derivatives at argument $x = \eta l \sin\theta$. We have

$$A_{\mathrm{x}l} = -\frac{i\, 2\pi q a}{cT} \frac{e^{ik|\mathbf{x}|}}{|\mathbf{x}|} J_l'(\eta l \sin\theta) \tag{11.29}$$

$$A_{\mathrm{y}l} = \frac{2\pi q a}{cT} \frac{e^{ik|\mathbf{x}|}}{|\mathbf{x}|} \frac{J_l(\eta l \sin\theta)}{\eta \sin\theta} \tag{11.30}$$

where we denoted $J_l'(x) = dJ_l/dx$. We are now ready to calculate the radiated power in point $\mathbf{x}$. According to Eq.(11.4) this is given by

$$\frac{dP}{d\Omega} = (\mathbf{S}\cdot\mathbf{n})(t)|\mathbf{x}|^2 = \frac{c}{4\pi} |\mathbf{x}|^2 (|\mathbf{E}|^2 - (\mathbf{E}\cdot\mathbf{n})^2) = \frac{c}{4\pi} |\mathbf{x}|^2 |\mathbf{E}|^2 \quad (|\mathbf{x}| \to \infty) \tag{11.31}$$

where we used Eq.(11.8). Note that here we now calculate the received power per solid angle $d\Omega$ in $\mathbf{x}$ directly at time t and do no need the factor κ which arose from the transformation to the retarded time t' in Eq.(11.4) . Now since

$$|\mathbf{B}| = |\mathbf{n} \times \mathbf{E}| = |\mathbf{E}|$$

we can write this also as

$$\frac{dP}{d\Omega} = \frac{c}{4\pi}|\mathbf{x}|^2|\mathbf{B}|^2 = \frac{c}{4\pi}|\mathbf{x}|^2|\nabla \times \mathbf{A}|^2 \tag{11.32}$$

where we used that $\mathbf{B} = \nabla \times \mathbf{A}$. Let us express this in terms of the Fourier coefficients $\mathbf{A}_l(\mathbf{x})$. We have

$$\nabla \times \mathbf{A}(\mathbf{x}, t) = \sum_{l=-\infty}^{+\infty} (\nabla \times \mathbf{A}_l(\mathbf{x}))\, e^{i\omega_0 l t}$$

and therefore

$$|\nabla \times \mathbf{A}(\mathbf{x}, t)|^2 = \sum_{l,m=-\infty}^{+\infty} (\nabla \times \mathbf{A}_l(\mathbf{x})) \cdot (\nabla \times \mathbf{A}_m(\mathbf{x}))\, e^{i\omega_0 (l+m)t}$$

Consequently

$$\begin{aligned}
\frac{1}{T}\int_0^T |\nabla \times \mathbf{A}(\mathbf{x}, t)|^2 &= \sum_{l,m=-\infty}^{+\infty} (\nabla \times \mathbf{A}_l(\mathbf{x})) \cdot (\nabla \times \mathbf{A}_m(\mathbf{x}))\, \delta_{l+m,0} \\
&= \sum_{l=-\infty}^{+\infty} (\nabla \times \mathbf{A}_l(\mathbf{x})) \cdot (\nabla \times \mathbf{A}_{-l}(\mathbf{x})) = \sum_{l=-\infty}^{+\infty} |\nabla \times \mathbf{A}_l(\mathbf{x})|^2
\end{aligned}$$

where we used that $\mathbf{A}_l^* = \mathbf{A}_{-l}$ and used the notation $|\mathbf{a}|^2 = \mathbf{a} \cdot \mathbf{a}^*$. From Eq.(11.32) we then see that the received power per solid angle in $\mathbf{x}$ is given by

$$\langle \frac{dP}{d\Omega} \rangle = \frac{c}{4\pi}|\mathbf{x}|^2 \sum_{l=-\infty}^{+\infty} |\nabla \times \mathbf{A}_l(\mathbf{x})|^2 \tag{11.33}$$

where the brackets denote the average over one period. We now only need to calculate the outer product in this expression. We have

$$\nabla \times \mathbf{A}_l = \begin{pmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \\ \frac{\partial}{\partial z} \end{pmatrix} \times \begin{pmatrix} A_{\mathrm{x}l} \\ A_{\mathrm{y}l} \\ 0 \end{pmatrix} = \begin{pmatrix} -\frac{\partial}{\partial z} A_{\mathrm{y}l} \\ \frac{\partial}{\partial z} A_{\mathrm{x}l} \\ \frac{\partial}{\partial x} A_{\mathrm{y}l} - \frac{\partial}{\partial y} A_{\mathrm{x}l} \end{pmatrix}$$

If we denote

$$\begin{aligned}
\alpha_{\mathrm{x}} &= -i\frac{\omega_0 a q}{c} J_l'(l\eta \sin\theta) \\
\alpha_{\mathrm{y}} &= \frac{\omega_0 a q}{c} \frac{J_l(l\eta \sin\theta)}{\eta \sin\theta}
\end{aligned}$$

then using

$$\frac{\partial}{\partial x_j}\frac{e^{ik|\mathbf{x}|}}{|\mathbf{x}|} = e^{ik|\mathbf{x}|}\left(\frac{ik}{|\mathbf{x}|} - \frac{1}{|\mathbf{x}|^2}\right)\frac{\partial |\mathbf{x}|}{\partial x_j} = x_j e^{ik|\mathbf{x}|}\left(\frac{ik}{|\mathbf{x}|^2} - \frac{1}{|\mathbf{x}|^3}\right) = ik\, x_j\, \frac{e^{ik|\mathbf{x}|}}{|\mathbf{x}|^2} \quad (|\mathbf{x}| \to \infty)$$

we have

$$\nabla \times \mathbf{A}_l = ik\frac{e^{ik|\mathbf{x}|}}{|\mathbf{x}|^2}\begin{pmatrix} -\alpha_{\mathrm{y}} z \\ \alpha_{\mathrm{x}} z \\ -\alpha_{\mathrm{x}} y \end{pmatrix}$$

where we need to remember that the observation point $\mathbf{x}$ was in the plane $x = 0$. We therefore find that

$$|\nabla \times \mathbf{A}_l|^2 = \frac{k^2}{|\mathbf{x}|^4}\left((|\alpha_{\mathrm{y}}|^2 + |\alpha_{\mathrm{x}}|^2)z^2 + |\alpha_{\mathrm{x}}|^2 y^2\right)$$

If we further use that $z = |\mathbf{x}|\cos\theta$ and $y = |\mathbf{x}|\sin\theta$ then this yields

$$\begin{aligned}|\nabla \times \mathbf{A}_l|^2 &= \frac{k^2}{|\mathbf{x}|^2}\left((|\alpha_{\mathrm{y}}|^2 + |\alpha_{\mathrm{x}}|^2)\cos^2\theta + |\alpha_{\mathrm{x}}|^2\sin^2\theta\right) = \frac{k^2}{|\mathbf{x}|^2}\left(|\alpha_{\mathrm{y}}|^2\cos^2\theta + |\alpha_{\mathrm{x}}|^2\right) \\ &= \left(\frac{\omega_0 akq}{c|\mathbf{x}|}\right)^2\left\{\frac{\cot^2\theta}{\eta^2}J_l^2(\eta l\sin\theta) + J_l'^2(\eta l\sin\theta)\right\} \qquad (11.34)\end{aligned}$$

If we recall that $\eta = \omega_0 a/c$ and $k = \omega_0 l/c$ we have from Eq.(11.33) that

$$\begin{aligned}\langle\frac{dP}{d\Omega}\rangle &= \frac{c}{4\pi}\left(\frac{\omega_0 q}{c}\right)^2\sum_{l=-\infty}^{+\infty} l^2\left\{\cot^2\theta J_l^2(\eta l\sin\theta) + \eta^2\, J_l'^2(\eta l\sin\theta)\right\} \\ &= \frac{\omega_0^2 q^2}{2\pi c}\sum_{l=1}^{\infty} l^2\left\{\cot^2\theta J_l^2(\eta l\sin\theta) + \eta^2\, J_l'^2(\eta l\sin\theta)\right\} = \sum_{l=0}^{\infty}\frac{dP_l}{d\Omega}\end{aligned}$$

where in the second step we used that the Bessel functions are even and we defined

$$\frac{dP_l}{d\Omega} = \frac{\omega_0^2 q^2}{2\pi c}\, l^2\,\left\{\cot^2\theta J_l^2(\eta l\sin\theta) + \eta^2\, J_l'^2(\eta l\sin\theta)\right\}$$

This expression gives the radiated power per solid angle and per cycle for each harmonic frequency $\omega_0 l$. The total radiated power is calculated from

$$\begin{aligned}P_l &= \int d\Omega\,\frac{dP_l}{d\Omega} = \int_0^{2\pi} d\phi\int_0^{\pi} d\theta\,\sin\theta\,\frac{dP_l}{d\Omega} \\ &= \frac{\omega_0^2 l^2 q^2}{c}\int_0^{\pi} d\theta\,\sin\theta\,\left\{\cot^2\theta J_l^2(\eta l\sin\theta) + \eta^2\, J_l'^2(\eta l\sin\theta)\right\}\end{aligned}$$

This integral an be manipulated using various special relations for the Bessel functions. We will not do this here and refer to Landau and Lishitz [6] for all details. Another reference to this calculation is [7]. The result is that in the relativistic limit $\gamma \gg 1$ the radiated power has the form

$$\begin{aligned}P_l &= 0.52\left(\frac{q^2}{2\pi c}\right)\omega_0^2\, l^{\frac{1}{3}} \qquad (1 \ll l \ll \gamma^3) \\ P_l &= \frac{1}{2\sqrt{\pi}}\left(\frac{q^2}{2\pi c}\right)\omega_0^2\left(\frac{l}{\gamma}\right)^{\frac{1}{2}} e^{-2l/3\gamma^3} \qquad (l \gg \gamma^3)\end{aligned}$$

which looks qualitatively as follows

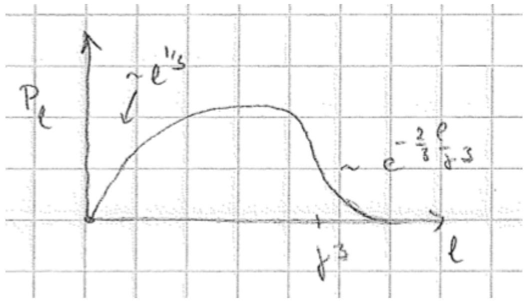

Figure 11.10: Distribution of intensities of radiation at frequency $\omega_0 l$.

Chapter 12

Radiation reaction

In this Chapter we will investigate the how energy and momentum loss to the radiation field creates a back reaction force on accelerated charged particles. This radiation reaction is describes by the Lorentz-Dirac equation. However, we will see that this equation has a number of unphysical features. We will look into some of the solutions of this equation and discuss some of its modifications.

12.1 Point charges and infinities

From our discussion in Chapter 6 we know that the motion of a charged particle is governed by the Lorentz force law.

$$m\frac{Dv^\mu}{d\tau} = \frac{q}{c}\sum_{\nu=0}^{3} F^{\mu\nu} v_\nu$$

where v^μ is the four-velocity of the particle. We also know that the particle will radiate as soon as it accelerates. This means that the particle will loose energy and momentum to the emitted radiation field. As a consequence of this energy-momentum loss there will be a reaction force back on the charged particle. This is described by the Lorentz force law if we include into $F^{\mu\nu}$ also the radiation fields produced by the charge itself. The force produced by these fields is called the *radiation reaction force* and the corresponding fields will be denoted by $F_r^{\mu\nu}$. The problem, however, is that these fields are singular at the position of the point particle and therefore yield an infinite reaction force. These were exactly the problems at the late 19th and early 20th century when Lorentz and others studied the electron. In the end it was found that usable equations for the reaction force can be derived provided the infinities are absorbed into a redefinition of the mass of the electron. Let us illustrate the idea with a simple example. Consider a simple point charge q and let us calculate the energy of its corresponding electromagnetic field. This is easiest in the rest frame of the particle, where we only have a spherically symmetric electric field. The energy density in this field is given by Eq.(10.8). Since we only have an electric field this gives the field energy

$$E = \frac{1}{8\pi}\int d\mathbf{x}|\mathbf{E}|^2 = \frac{1}{8\pi}\int d\mathbf{x}\left(\frac{q}{|\mathbf{x}|^2}\right)^2 = \lim_{r_0\to 0}\frac{q^2}{8\pi}\int_{r_0}^{\infty} dr\, 4\pi\frac{1}{r^2} = \lim_{r_0\to 0}\frac{q^2}{2r_0}$$

where we put a lower limit r_0 on the radial integral to show the way in which the energy approaches infinity. Let us assume as Lorentz and others did that the charge describes the electron, which is a fundamental physical object. Then the we could imagine that it has an internal structure and an extension characterized by the radius r_0. We can then according to the formula $E = mc^2$ assign an electromagnetic mass m_e to the electron given by

$$E = m_e c^2 = \frac{q^2}{2r_0} \tag{12.1}$$

If we assume all mass is of electromagnetic origin then for the electron we find $r_0 = 1.5 \times 10^{-15}$m. The infinities can therefore be avoided by assuming a finite extension. The physical question is then whether the internal structure of the particle is relevant for radiation reaction. To study this Lorentz considered a simple model for the electron.

12.2 The Lorentz model

To describe the effect of radiation reaction on the motion of the electron Lorentz [8, 9] considered a hollow sphere with a uniform surface charge density. If the electron is at rest the forces on each piece of the charge balance each other (assuming that electric forces are balanced by unknown internal forces).

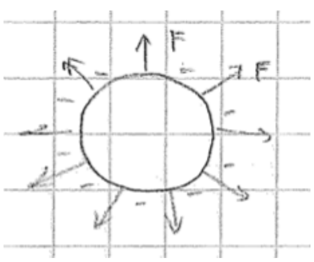

Figure 12.1: Charged sphere at rest.

However, when the sphere moves in a non-uniform manner the electric forces no longer balance because the electromagnetic fields take times to travel from one pice of the sphere to another.

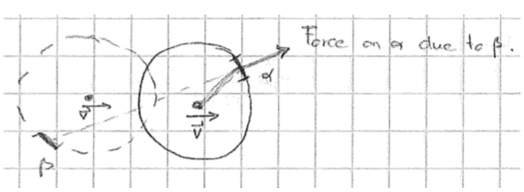

Figure 12.2: Charged sphere in motion.

For instance, the force on a piece α due to a piece β depends on the position of β at an earlier time. Lorentz then chose a reference frame in which the particle was instantaneously at rest at time t, i.e. $\dot{\mathbf{z}}(t) = 0$ for a charge element at position $\mathbf{z}(t)$, and assumed that none of the quantities $\dot{\mathbf{z}}, \ddot{\mathbf{z}}, \dddot{\mathbf{z}}$ changes very much during the time it takes for an electromagnetic signal to cross the electron, such that they can be expanded in a Taylor series around the retarded time

$$\mathbf{z}(t') = \mathbf{z}(t - \frac{R}{c}) = \mathbf{z}(t) - \frac{R}{c}\dot{\mathbf{z}}(t) + \frac{1}{2}\left(\frac{R}{c}\right)^2 \ddot{\mathbf{z}}(t) + \dots$$

where $R = |\mathbf{x} - \mathbf{z}(t')|$ is the distance form $\mathbf{x}$ to the charge element at $\mathbf{z}(t')$. By expanding the various contributions of the electric field and performing an integral of the sphere Lorentz found

[8, 9] that the net force $\mathbf{F}$ due to the self-force is given by

$$\mathbf{F}_r = \frac{2q^2}{3c^3}\dddot{\mathbf{z}}(t) - \frac{2q^2}{3r_0c^2}\ddot{\mathbf{z}}(t) = \frac{2q^2}{3c^3}\dddot{\mathbf{z}}(t) - m_e\ddot{\mathbf{z}}(t) \tag{12.2}$$

where $\mathbf{z}(t)$ describes the position of the center of the sphere and q is the total charge of the sphere. There are higher order terms as well, but these vanish in the limit that $r_0 \to \infty$. Here we further define the electromagnetic mass by $m_e = 2q^3/3r_0c^2$. This differs from the previous choice by a factor $2/3$ but this is due to the specific model that we use. Note that again $m_e \to \infty$ for $r_0 \to 0$. The total force on the particle is then given by the sum of the external and the reaction force. This is given by

$$m_0\ddot{\mathbf{z}}(t) = \mathbf{F}_{\text{ext}} + \frac{2q^2}{3c^3}\dddot{\mathbf{z}}(t) - m_e\ddot{\mathbf{z}}(t)$$

or equivalently

$$m\ddot{\mathbf{z}}(t) = \mathbf{F}_{\text{ext}} + \frac{2q^2}{3c^3}\dddot{\mathbf{z}}(t) \tag{12.3}$$

where we defined the physical mass $m = m_0 + m_e$. The mass m_0 is an unobservable mass parameter. Eq.(12.3) is known as the Lorentz equation. Somewhat unusually it is a third order differential equation. One therefore needs to specify an initial position, velocity and acceleration to solve it.

12.3 The Lorentz-Dirac equation

The Lorentz equation in the previous section was derived in a specific reference frame by expansion in powers of R/c where R is essentially the radius of the electron. However, the last term in Eq.(12.3) is independent of this radius and the higher order terms in $1/c^n$ $(n \geq 4)$ vanish when $R \to 0$. Therefore Eq.(12.3) has the potential of being exact and it will be worthwhile to find its relativistic generalization. The obvious guess in Minkowskian coordinates is

$$m\frac{d^2z^\mu}{d\tau^2} = F^\mu_{\text{ext}} + \frac{2q^2}{3c^3}\frac{d^3z^\mu}{d\tau^3}$$

We should, however, remember that the Lorentz equation was derived in the instantaneous rest frame of the electron and therefore in a general Lorentz frame terms proportional to $dz^\mu/d\tau$ may appear. We therefore make the Ansatz

$$m\frac{d^2z^\mu}{d\tau^2} = F^\mu_{\text{ext}} + \frac{2q^2}{3c^3}\left(\frac{d^3z^\mu}{d\tau^3} + S\frac{dz^\mu}{d\tau}\right) \tag{12.4}$$

where S is a scalar function to be determined. It is readily derived that S can not be equal to zero. From the condition (6.9)

$$\sum_{\mu=0}^{3} v_\mu v^\mu = -c^2 \tag{12.5}$$

where $v^\mu = dz^\mu/d\tau$ we derived Eq.(6.12) which in Minkowskian coordinates reads

$$0 = \sum_{\mu=0}^{3} v_\mu \frac{dv^\mu}{d\tau} \tag{12.6}$$

and which in turn yields

$$0 = \sum_{\mu=0}^{3} v_\mu F^\mu \tag{12.7}$$

where F is any applied force assumed to be proportional to the four-momentum. When we apply this condition to both sides of Eq.(12.4) we obtain the relation

$$0 = \sum_{\mu=0}^{3} \left(\frac{dz_\mu}{d\tau} \frac{d^3 z^\mu}{d\tau^3} + S \frac{dz_\mu}{d\tau} \frac{dz^\mu}{d\tau} \right) = \sum_{\mu=0}^{3} \frac{dz_\mu}{d\tau} \frac{d^3 z^\mu}{d\tau^3} - c^2 S$$

and consequently we find that

$$S = \frac{1}{c^2} \sum_{\mu=0}^{3} \frac{dz_\mu}{d\tau} \frac{d^3 z^\mu}{d\tau^3}$$

This equation can be rewritten since by differentiation of Eq.(12.6) with respect to the proper time τ we have

$$0 = \sum_{\mu=0}^{3} \frac{dv_\mu}{d\tau} \frac{dv^\mu}{d\tau} + \sum_{\mu=0}^{3} v_\mu \frac{d^2 v^\mu}{d\tau^2} \tag{12.8}$$

from which it follows that

$$S = -\frac{1}{c^2} \sum_{\mu=0}^{3} \frac{d^2 z_\mu}{d\tau^2} \frac{d^2 z^\mu}{d\tau^2} = -\frac{1}{c^2} \sum_{\mu=0}^{3} \frac{dv_\mu}{d\tau} \frac{dv^\mu}{d\tau} \tag{12.9}$$

Our relativistic proposal for the radiation reaction equation (12.4) therefore becomes

$$m \frac{dv^\mu}{d\tau} = F^\mu_{\mathrm{ext}} + \frac{2q^2}{3c^3} \left(\frac{d^2 v^\mu}{d\tau^2} - \frac{1}{c^2} \left[\sum_{\nu=0}^{3} \frac{dv_\nu}{d\tau} \frac{dv^\nu}{d\tau} \right] v^\mu \right) \tag{12.10}$$

This equation is known as the *Lorentz-Dirac equation*. Dirac derived this equation in a very different way than Lorentz using point particles from the start and conservation of energy-momentum. A good discussion of this derivation is given in [10, 11]. Before we discuss some of the solutions of the Lorentz-Dirac equation let us first give some physical meaning to the extra term in Eq.(12.10). We first write it as

$$\sum_{\nu=0}^{3} \frac{dv_\nu}{d\tau} \frac{dv^\nu}{d\tau} = \sum_{\nu=0}^{3} \frac{dv_\nu}{dt} \frac{dv^\nu}{dt} \left(\frac{dt}{d\tau} \right)^2 = \gamma^2 \sum_{\nu=0}^{3} \frac{dv_\nu}{dt} \frac{dv^\nu}{dt}$$

where we reparametrized the motion $z^\mu(\tau)$ of the particle by the time t in the Lorentz frame from which we observe the particle and used that $dt/d\tau = (1 - v^2/c^2)^{-1/2} = \gamma$. Let us now calculate the last term more explicitly. the four-velocity has the form (see Eq.(6.17))

$$v = \gamma(c, \mathbf{v})$$

and therefore

$$\frac{dv}{dt} = (c\frac{d\gamma}{dt}, \frac{d\gamma}{dt}\mathbf{v} + \gamma \frac{d\mathbf{v}}{dt})$$

Since

$$\frac{d\gamma}{dt} = \frac{d}{dt} \left(1 - \frac{\mathbf{v}^2}{c^2} \right)^{-\frac{1}{2}} = \mathbf{v} \cdot \frac{d\mathbf{v}}{dt} \frac{1}{c^2} \gamma^3$$

we have

$$\frac{dv}{dt} = (\frac{1}{c}\gamma^3 \mathbf{v} \cdot \frac{d\mathbf{v}}{dt}, \mathbf{v} \cdot \frac{d\mathbf{v}}{dt} \frac{\mathbf{v}}{c^2} \gamma^3 + \gamma \frac{d\mathbf{v}}{dt})$$

from which it follows that

$$\begin{aligned}
\sum_{\nu=0}^{3} \frac{dv_\nu}{dt}\frac{dv^\nu}{dt} &= -\frac{\gamma^6}{c^2}\left(\mathbf{v}\cdot\frac{d\mathbf{v}}{dt}\right)^2 + \left(\mathbf{v}\cdot\frac{d\mathbf{v}}{dt}\frac{\mathbf{v}}{c^2}\gamma^3 + \gamma\frac{d\mathbf{v}}{dt}\right)^2 \\
&= -\frac{\gamma^6}{c^2}\left(1-\frac{\mathbf{v}^2}{c^2}\right)\left(\mathbf{v}\cdot\frac{d\mathbf{v}}{dt}\right)^2 + \frac{2\gamma^4}{c^2}\left(\mathbf{v}\cdot\frac{d\mathbf{v}}{dt}\right)^2 + \gamma^2\left(\frac{d\mathbf{v}}{dt}\right)^2 \\
&= \gamma^2\left(\frac{\gamma^2}{c^2}\left(\mathbf{v}\cdot\frac{d\mathbf{v}}{dt}\right)^2 + \left(\frac{d\mathbf{v}}{dt}\right)^2\right) \\
&= \gamma^4\left(\left(\frac{d\mathbf{v}}{dt}\right)^2 + \frac{1}{c^2}\left[\left(\mathbf{v}\cdot\frac{d\mathbf{v}}{dt}\right)^2 - \mathbf{v}^2\left(\frac{d\mathbf{v}}{dt}\right)^2\right]\right) \\
&= \gamma^4\left(\left(\frac{d\mathbf{v}}{dt}\right)^2 - \frac{1}{c^2}\left(\mathbf{v}\times\frac{d\mathbf{v}}{dt}\right)^2\right) = c^2\gamma^4(\dot{\boldsymbol{\beta}}^2 - (\boldsymbol{\beta}\times\dot{\boldsymbol{\beta}})^2)
\end{aligned}$$

where $\boldsymbol{\beta} = \mathbf{v}/c$. We therefore find that

$$\sum_{\nu=0}^{3} \frac{dv_\nu}{d\tau}\frac{dv^\nu}{d\tau} = c^2\gamma^6(\dot{\boldsymbol{\beta}}^2 - (\boldsymbol{\beta}\times\dot{\boldsymbol{\beta}})^2)$$

If we compare this expression to the formula of the total radiated power P of Eq.(11.18) we see that

$$P = \frac{2q^2}{3c}\frac{1}{(1-\beta^2)^3}(\dot{\boldsymbol{\beta}}^2 - (\boldsymbol{\beta}\times\dot{\boldsymbol{\beta}})^2) = \frac{2q^2}{3c^3}\sum_{\nu=0}^{3}\frac{dv_\nu}{d\tau}\frac{dv^\nu}{d\tau} \tag{12.11}$$

which gives a explicitly Lorentz invariant description of the radiated power. We can therefore equivalently write Eq.(12.10) as

$$m\frac{dv^\mu}{d\tau} = F^\mu_{\rm ext} + \frac{2q^2}{3c^3}\frac{d^2v^\mu}{d\tau^2} - \frac{P}{c^2}v^\mu$$

Let us now take the external force to be given by an external electromagnetic field $F^{\mu\nu}_{\rm ext}$. Then we can write Eq.(12.10) as

$$m\frac{dv^\mu}{d\tau} = \frac{q}{c}\sum_{\nu=0}^{3}F^{\mu\nu}_{\rm ext}v_\nu + \frac{2q^2}{3c^3}\left(\frac{d^2v^\mu}{d\tau^2} - \frac{1}{c^2}\left[\sum_{\nu=0}^{3}\frac{dv_\nu}{d\tau}\frac{dv^\nu}{d\tau}\right]v^\mu\right) \tag{12.12}$$

Also the last term can be written in terms of a reaction force tensor $F^{\mu\nu}_r$ such that

$$m\frac{dv^\mu}{d\tau} = \frac{q}{c}\sum_{\nu=0}^{3}(F^{\mu\nu}_{\rm ext} + F^{\mu\nu}_r)v_\nu \tag{12.13}$$

where

$$F^{\mu\nu}_r = \frac{2q}{3c^4}\left(v^\mu\frac{d^2v^\nu}{d\tau^2} - v^\nu\frac{d^2v^\mu}{d\tau^2}\right)$$

This is readily checked. We have

$$\begin{aligned}\frac{q}{c}\sum_{\nu=0}^{3} F_r^{\mu\nu} v_\nu &= \frac{2q^2}{3c^5}\sum_{\nu=0}^{3}\left(v^\mu \frac{d^2 v^\nu}{d\tau^2} v_\nu - v^\nu v_\nu \frac{d^2 v^\mu}{d\tau^2}\right) \\ &= \frac{2q^2}{3c^5}\left(-v^\mu \sum_{\nu=0}^{3}\frac{dv_\nu}{d\tau}\frac{dv^\nu}{d\tau} + c^2 \frac{d^2 v^\mu}{d\tau^2}\right) \\ &= \frac{2q^2}{3c^3}\left(\frac{d^2 v^\mu}{d\tau^2} - \frac{1}{c^2}\sum_{\nu=0}^{3}\frac{dv_\nu}{d\tau}\frac{dv^\nu}{d\tau}\right)\end{aligned}$$

which is exactly the last term in Eq.(12.12). The Lorentz-Dirac equation can therefore be written in the form of Eq.(12.13) which is exactly the Lorentz force law in which the electromagnetic field tensor now includes a piece $F_r^{\mu\nu}$ due to the fields produced by the charge itself. Note to now our logics has made a circle. In Chapter 6 we assumed that the derivative of the momentum was proportional to the four-velocity from which we concluded that they were connected by an anti-symmetric tensor $F^{\mu\nu}$. We then in Chapter 7 constructed the Maxwell equations for the electromagnetic tensor in terms of the four-current. By solving the Maxwell equations in Chapter 8 we could in Chapters 9 and 11 derive the Liénard-Wiechert potentials for the fields produced by moving charges. Using the Poynting vector from the energy-momentum tensor we then calculated the radiated power by moving particles. We then concluded that since the particle looses energy and momentum to the emitted field there must be a reaction force back on the particle. Here we find that this reaction force is described by the reaction field tensor $F_r^{\mu\nu}$. But now we see that due to $F_r^{\mu\nu}$ the change in momentum in Eq.(12.13) is not anymore linear in the four-velocity since $F_r^{\mu\nu}$ depends in a non-linear way on v^μ. The whole dynamics of a system of charged particles is still described by our fundamental Eqs.(7.25)-(7.27) What we have effectively done is to solve Eqs.(7.25) and (7.26) in terms of the four-current and put the solution into Eq.(7.27) thereby making the Lorentz force law a non-linear equation for the four-velocity. The coupling of several linear equations therefore leads to rather complicated nonlinear dynamics.

12.4 Solutions of the Lorentz-Dirac equation

12.4.1 Linear motion

Let us no study some solutions of the Lorentz-Dirac equation. As we will see we will run into a number of conceptual issues, which we will address at the end of this section. Let us consider Eq.(12.10) for the case of linear one-dimensional motion. Because of the condition (12.5) we can always write for motion along the x-axis that the four-velocity v^μ can be parametrized as

$$v = (c\cosh w(\tau), c\sinh w(\tau), 0, 0) \tag{12.14}$$

since then

$$\sum_{\mu=0}^{3} v_\mu v^\mu = -c^2(\cosh^2 w(\tau) - \sinh^2 w(\tau)) = -c^2$$

If we consider the four-velocity in a given Lorentz frame then we have (see Eq.(6.17)) that

$$v = \gamma(c, v_{\mathrm{x}}, 0, 0) \tag{12.15}$$

where v_{x} is the velocity along the x-axis. By comparison to Eq.(12.14) we can therefore identify

$$\begin{aligned}\gamma &= \cosh w(\tau) \\ v_{\mathrm{x}} &= c\tanh w(\tau)\end{aligned} \tag{12.16}$$

We further have that
$$\frac{dv}{d\tau} = \frac{dw}{d\tau}(c\sinh w(\tau), c\cosh w(\tau), 0, 0)$$
and therefore
$$\sum_{\nu=0}^{3} \frac{dv_\nu}{d\tau}\frac{dv^\nu}{d\tau} = c^2\left(\frac{dw}{d\tau}\right)^2(-\sinh^2 w(\tau) + \cosh^2 w(\tau)) = c^2\left(\frac{dw}{d\tau}\right)^2$$
Finally we have that
$$\frac{d^2v}{d\tau^2} = \frac{d^2w}{d\tau^2}(c\sinh w(\tau), c\cosh w(\tau), 0, 0) + \left(\frac{dw}{d\tau}\right)^2(c\cosh w(\tau), c\sinh w(\tau), 0, 0)$$
If we now insert all this information back into Eq.(12.10) we obtain the equation
$$mc\frac{dw}{d\tau}\begin{pmatrix}\sinh w\\ \cosh w\end{pmatrix} = \begin{pmatrix}F^0\\ F^1\end{pmatrix} + \frac{2q^2}{3c^3}\frac{d^2w}{d\tau^2}c\begin{pmatrix}\sinh w\\ \cosh w\end{pmatrix} \tag{12.17}$$
where we assume a force only acting along the x-axis. Now the condition (12.7) on the external force gives the relation
$$F^0v_0 + F^1v_1 = 0 \implies F^0 = -\frac{v_1}{v_0}F^1 = \tanh w(\tau)F^1$$
If we further define the function $F(\tau)$ by
$$F^1(\tau) = \cosh w(\tau)F(\tau)$$
then we can write
$$F = (F^0, F^1, 0, 0) = (\sinh w(\tau)F(\tau), \cosh w(\tau)F(\tau), 0, 0)$$
and Eq.(12.17) attains the simple form
$$\frac{dw}{d\tau} - \frac{2q^2}{3mc^3}\frac{d^2w}{d\tau^2} = \frac{1}{mc}F(\tau)$$
If we further denote
$$\begin{aligned}\tau_0 &= \frac{2q^2}{3mc^3}\\ f(\tau) &= \frac{1}{mc}F(\tau)\end{aligned} \tag{12.18}$$
then we obtain the simple linear differential equation
$$\frac{dw}{d\tau} - \tau_0\frac{d^2w}{d\tau^2} = f(\tau) \tag{12.19}$$
where the constant τ_0 has the dimension of time. It is not difficult to find a general solution of this equation. If we write
$$h(\tau) = \frac{dw}{dt}$$
The equation is transformed to the first order equation
$$h(\tau) - \tau_0\frac{dh}{d\tau} = f(\tau)$$

We apply the standard method of variation of constants for solving this equation. We first consider the homogeneous solution $f = 0$. In this case the general solution is obviously

$$h(\tau) = A\, e^{\frac{\tau}{\tau_0}}$$

This is clearly an unphysical solution unless $A = 0$ as the particle will accelerate exponentially in time in the absence of any force. We therefore need to take $A = 0$ for the homogeneous solution. The remaining particular solution is obtained by variation of constants. We say

$$h(\tau) = e^{\frac{\tau}{\tau_0}}\, g(\tau)$$

which inserted back into the differential equation gives

$$-\tau_0\, e^{\frac{\tau}{\tau_0}}\, \frac{dg}{d\tau} = f(\tau)$$

and which is easily integrated to give

$$g(\tau) = g(\sigma) - \frac{1}{\tau_0} \int_\sigma^\tau d\tau'\, e^{-\frac{\tau'}{\tau_0}}\, f(\tau')$$

We therefore find that

$$h(\tau) = e^{\frac{\tau}{\tau_0}} \left[g(\sigma) - \frac{1}{\tau_0} \int_\sigma^\tau d\tau'\, e^{-\frac{\tau'}{\tau_0}}\, f(\tau') \right]$$

where $g(\sigma)$ must be determined by a boundary condition. We see that $h(\tau) \to \infty$ for $\tau \to \infty$ for any force profile unless

$$0 = \lim_{\tau \to \infty} \left[g(\sigma) - \frac{1}{\tau_0} \int_\sigma^\tau d\tau'\, e^{-\frac{\tau'}{\tau_0}}\, f(\tau') \right] \tag{12.20}$$

To prevent again unphysical run-away solutions we have to choose

$$g(\sigma) = \frac{1}{\tau_0} \int_\sigma^\infty d\tau'\, e^{-\frac{\tau'}{\tau_0}}\, f(\tau')$$

and we find that

$$h(\tau) = \frac{e^{\frac{\tau}{\tau_0}}}{\tau_0} \int_\tau^\infty d\tau'\, e^{-\frac{\tau'}{\tau_0}}\, f(\tau') = \frac{dw}{d\tau} \tag{12.21}$$

A further integration then gives

$$w(\tau) = w(a) + \frac{1}{\tau_0} \int_a^\tau ds\, e^{\frac{s}{\tau_0}} \int_s^\infty d\tau'\, e^{-\frac{\tau'}{\tau_0}}\, f(\tau')$$

where $w(a)$ is an initial value. Let us apply this to a simple case. We take

$$f(\tau) = \lambda\, \delta(\tau)$$

i.e. we apply a delta pulse of strength λ (if you prefer a more physically force you could approximate this with a sharp Gaussian function for instance). If we insert this force into Eq.(12.21) we find

$$\frac{dw}{d\tau} = \frac{e^{\frac{\tau}{\tau_0}}}{\tau_0} \int_\tau^\infty d\tau'\, e^{-\frac{\tau'}{\tau_0}}\, \lambda\, \delta(\tau') = \begin{cases} \frac{\lambda}{\tau_0}\, e^{\frac{\tau}{\tau_0}} & (\tau < 0) \\ 0 & (\tau > 0) \end{cases}$$

and therefore

$$w(\tau) = \begin{cases} w(-\infty) + \lambda\, e^{\frac{\tau}{\tau_0}} & (\tau < 0) \\ w(-\infty) + \lambda & (\tau > 0) \end{cases}$$

where the integration constants for $\tau > 0$ and $\tau < 0$ are connected by requiring $w(\tau)$ to be continuous for $\tau = 0$. If, for simplicity, we say that the particle has zero velocity for $\tau \to -\infty$ then we have the following situation

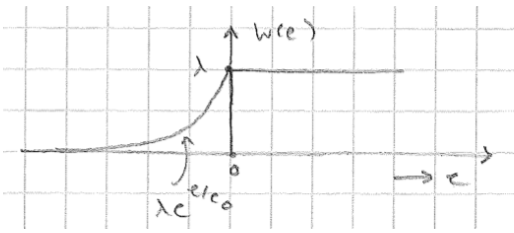

Figure 12.3: Solution for w for a delta pulse force.

The velocity of the particle is given by Eq.(12.16). We therefore observe the strange situation that the particle starts to move before the delta pulse at $\tau = 0$ is applied. This phenomena has been called pre-acceleration and is a strange feature of the Lorentz-Dirac equation. Note that the time-scale τ_0 on which this happens is very small. For the case of an electron

$$\tau_0 = 0.62 \times 10^{-23}\,\mathrm{s}$$

The phenomenon could therefore signify the breakdown of a classical regime and indicate the necessity of a quantum description of the problem. For a discussion see [10].

12.4.2 Harmonic motion and radiation damping

Let us finally discuss the following physical phenomenon Imagine a charge in harmonic motion, like a pendulum or a spring

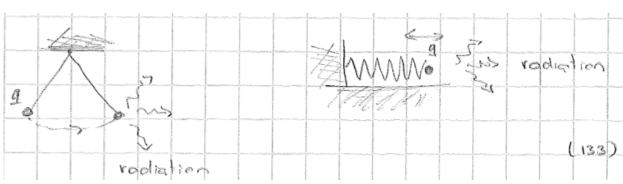

Figure 12.4: Harmonic motion with radiation damping.

If we let the charge oscillate then the particle accelerates and starts to radiate. The particle looses energy to the radiation field and consequently the oscillatory motion will be damped by the reaction force. This phenomenon is called *radiation damping*. Let us see how the Lorentz-Dirac equation describes this phenomenon. We consider one-dimension motion again and take for $F(\tau)$ in Eq.(12.18) the expression

$$F(\tau) = -k\,c \int_a^\tau d\tau'\, w(\tau') \tag{12.22}$$

where k is a spring constant and a an initial time. This equation represents a relativistic generalization of the harmonic force equation. Let us check this. In the non-relativistic limit for

$v_x \ll c$ we have according to Eq.(12.16) that

$$w = \tanh^{-1} \frac{v_{\mathbf{x}}}{c} \approx \frac{v_x}{c} = \frac{1}{c}\frac{dx}{d\tau} \tag{12.23}$$

in which limit we can also equalize the proper time τ and the time t in the Lorentz frame. So the non-relativistic limit becomes

$$F(\tau) = -k \int_a^\tau d\tau' \, \frac{dx}{d\tau'} = -k(x(\tau) - x(a)) \tag{12.24}$$

which is the usual harmonic force law. We may imagine other relativistic generalizations of this force law than (12.22) but this one is one of the simplest and nicely illustrates the phenomenon of radiation damping. With this choice of the force Eq.(12.19) becomes

$$\frac{dw}{d\tau} - \tau_0 \frac{d^2 w}{d\tau^2} = -\frac{k}{m} \int_a^\tau d\tau' \, w(\tau')$$

A subsequent differentiation of this equation then gives

$$\frac{d^2 w}{d\tau^2} - \tau_0 \frac{d^3 w}{d\tau^3} + \frac{k}{m} w = 0$$

If we denote by $\omega_0 = \sqrt{k/m}$ the harmonic frequency and use the Ansatz $w = e^{-i\alpha}$ then we find that

$$-\alpha^2 - i\tau_0 \, \alpha^3 + \omega_0^2 = 0$$

The three solutions to this equation are given by

$$\alpha_1 = \omega - i\gamma, \quad \alpha_2 = -\omega - i\gamma, \quad \alpha_3 = i(\frac{1}{\tau_0} + 2\gamma)$$

where

$$\begin{aligned}
\omega &= \frac{1}{2}\sqrt{3}\,(a_+ - a_-)\frac{1}{\tau_0} \\
\gamma &= \frac{1}{\tau_0}\left[\frac{1}{2}(a_+ + a_-) - \frac{1}{3}\right] \\
a_\pm &= \left[\frac{(\omega_0\tau_0)^2}{2} + \frac{1}{27} \pm \sqrt{\left(\frac{(\omega_0\tau_0)^2}{2} + \frac{1}{27}\right)^2 - \left(\frac{1}{27}\right)^2}\,\right]^{1/3} = \frac{1}{3} \pm \frac{\omega_0\tau_0}{\sqrt{3}} + \frac{(\omega_0\tau_0)^2}{2} + \ldots
\end{aligned}$$

These equations imply that $\omega, \gamma \geq 0$. The general solution for w is therefore given by

$$\begin{aligned}
w(\tau) &= A_1 \, e^{-i\alpha_1 \tau} + A_2 \, e^{-i\alpha_2 \tau} + A_3 \, e^{-i\alpha_3 \tau} \\
&= e^{-\gamma\tau}(A_1 \, e^{-i\omega \tau} + A_2 \, e^{i\omega \tau}) + A_3 \, e^{(\frac{1}{\tau_0}+2\gamma)\,\tau}
\end{aligned}$$

In order for $w(\tau)$ to remain finite for $\tau \to \infty$ we must have $A_3 = 0$. The general solution can therefore be written as

$$w(\tau) = A \, e^{-\gamma\tau} \, \sin(\omega\tau + \phi)$$

where the amplitude A and phase ϕ are determined by the initial conditions on w and $dw/d\tau$. We therefore observe a damped harmonic motion with damping determined by γ. In the limit $\omega_0 \ll 1/\tau_0$ it follows that

$$\omega \approx \omega_0 \quad , \quad \gamma \approx \frac{1}{2}\omega_0^2 \tau_0$$

and therefore

$$w(\tau) = A\, e^{-\frac{1}{2}\omega_0^2 \tau_0 \tau} \, \sin(\omega_0 \tau + \phi)$$

For $A \ll 1 (v_{\rm x} \ll c)$ we have $v_{\rm x} \approx cw$. The frequency $1/\tau_0$ is of the order of 10^{23} Hz for a single electron and therefore the approximation $\omega_0 \tau_0 \ll 1$ is a very realistic one for most applications of electromagnetic theory. The damping rate is therefore in general very small. However, radiation reaction effects have recently been under active study due to the study of electronic motion in ultra-intense laser fields [12, 13, 14].

12.5 The Landau-Lifshitz equation

As we have seen the Lorentz-Dirac equation has a number of artificial properties, such as run-away solutions and pre-acceleration. These undesirable acausal features of the equation can be traced back to the fact that it contains third order time-derivatives derivatives. One way to turn the equation into a lower order equation which does not possess any causality probelms been indicated by Landau and Lifshitz. Let us first write Eq.(12.12) as

$$\frac{dv^\mu}{d\tau} = \frac{q}{mc} \sum_{\nu=0}^{3} F^{\mu\nu}_{\rm ext} v_\nu + \tau_0 \left(\frac{d^2 v^\mu}{d\tau^2} - \frac{1}{c^2} \left[\sum_{\nu=0}^{3} \frac{dv_\nu}{d\tau} \frac{dv^\nu}{d\tau} \right] v^\mu \right) \tag{12.25}$$

Since τ_0 is a very small number we may decide to expand the solution in powers of τ_0. The zeroth order approximation is simply given by

$$\frac{dv^\mu}{d\tau} = \frac{q}{mc} \sum_{\nu=0}^{3} F^{\mu\nu}_{\rm ext} v_\nu$$

from which we by differentiation obtain

$$\begin{aligned} \frac{d^2 v^\mu}{d\tau^2} &= \frac{q}{mc} \sum_{\nu=0}^{3} \left(\frac{dF^{\mu\nu}_{\rm ext}}{d\tau} v_\nu + F^{\mu\nu}_{\rm ext} \frac{dv_\nu}{d\tau} \right) = \frac{q}{mc} \sum_{\nu,\rho=0}^{3} \left(\frac{dF^{\mu\nu}_{\rm ext}}{dx^\rho} \frac{dx^\rho}{d\tau} v_\nu + F^{\mu\nu}_{\rm ext} \frac{q}{mc} \sum_{\eta=0}^{3} F_{\nu\eta,{\rm ext}} v^\eta \right) \\ &= \frac{q}{mc} \sum_{\nu,\rho} \frac{dF^{\mu\nu}_{\rm ext}}{dx^\rho} v^\rho v_\nu + \left(\frac{q}{mc} \right)^2 \sum_{\nu,\eta} F^{\mu\nu}_{\rm ext} F_{\nu\eta,{\rm ext}} v^\eta \end{aligned}$$

If we insert all these expressions back into Eq.(12.25) we obtain

$$\begin{aligned} \frac{dv^\mu}{d\tau} &= \frac{q}{mc} \sum_{\nu=0}^{3} F^{\mu\nu}_{\rm ext} v_\nu + \\ &\tau_0 \left(\frac{q}{mc} \sum_{\nu,\rho} \frac{dF^{\mu\nu}_{\rm ext}}{dx^\rho} v^\rho v_\nu + \left(\frac{q}{mc} \right)^2 \sum_{\nu,\eta} F^{\mu\nu}_{\rm ext} F_{\nu\eta,{\rm ext}} v^\eta - \frac{1}{c^2} \left(\frac{q}{mc} \right)^2 \sum_{\nu,\rho,\eta} F_{\nu\rho,{\rm ext}} F^{\nu\eta}_{\rm ext} v^\rho v_\eta v^\mu \right) \end{aligned}$$

in which we neglected terms of order τ_0^2. Using Eq.(12.8) we also could have written this equation in alternative ways which would only matter to second order in τ_0. This equation is known as the Landau-Lifshitz equation [6]. However, the most important thing is that the right hand side of the equation only depends on velocities and not on their time-derivatives and therefore cures the a-causal features of the Lorentz-Dirac equation. We may wonder how fundamental this equation is as we neglected higher order terms in τ_0. We should, however, keep in mind that these classical equations are anyway only approximate as we, for instance, neglected quantum effects. In any case both the Landau-Lifshitz and Lorentz-Dirac equation are subject of current investigations for electronic motion in intense laser fields [14].

Chapter 13

Maxwell action and gauge curvature

13.1 Integration of differential forms

13.1.1 Differential forms and volume

13.1.2 Stokes' equation

13.1.3 A differential form inner product

13.2 Maxwell action

13.3 Gauge curvature

Appendix A

Diagonalization of non-degenerate metric

Appendix B

Covariant divergence from exterior calculus

We give here a proof of Eq.(7.61). To do this we have to evaluate $\star d \star$ on a general p-form ω. Let us first derive a useful formula. From Eq.(3.107) we see that we can write

$$\epsilon^{i_1 \ldots i_n} = \frac{1}{g}\, \epsilon_{i_1 \ldots i_n} \tag{B.1}$$

We will use this formula below. Let us take an arbitrary p-form ω

$$\omega = \sum_{i_1 < \ldots < i_p}^{n} \omega_{i_1 \ldots i_p}\, dx^{i_1} \wedge \ldots \wedge dx^{i_p}$$

Then according to Eq.(3.101) we have

$$\star\, \omega = \frac{1}{(n-p)!p!} \sum_{j_1 \ldots j_p, i_{p+1} \ldots i_n}^{n} \sqrt{|g|}\, \omega^{j_1 \ldots j_p} \epsilon_{j_1 \ldots j_p i_{p+1} \ldots i_n} dx^{i_{p+1}} \wedge \ldots \wedge dx^{i_n}$$

If we now apply the exterior derivative to this expression we find

$$\begin{aligned} d \star \omega &= \frac{1}{(n-p)!p!} \sum_{j_1 \ldots j_p, i_{p+1} \ldots i_n}^{n} \frac{\partial}{\partial x^k} \left(\sqrt{|g|}\, \omega^{j_1 \ldots j_p} \right) \epsilon_{j_1 \ldots j_p i_{p+1} \ldots i_n} dx^k \wedge dx^{i_{p+1}} \wedge \ldots \wedge dx^{i_n} \\ &= \frac{1}{p!(n-p)!(n-p+1)!} \sum_{\substack{k, j_1 \ldots j_p, i_{p+1} \ldots i_n \\ l_1 \ldots l_{n-p+1}}}^{n} \delta^{l_1 \ldots l_{n-p+1}}_{k\, i_{p+1} \ldots i_n} \epsilon_{j_1 \ldots j_p i_{p+1} \ldots i_n} \frac{\partial}{\partial x^k} \left(\sqrt{|g|}\, \omega^{j_1 \ldots j_p} \right) \\ &\quad \times dx^{l_1} \wedge \ldots \wedge dx^{l_{n-p+1}} \\ &= \frac{1}{(n-p+1)!} \sum_{l_1 \ldots l_{n-p+1}}^{n} \eta_{l_1 \ldots l_{n-p+1}}\, dx^{l_1} \wedge \ldots \wedge dx^{l_{n-p+1}} \end{aligned}$$

where the coefficients of the $(n-p+1)$-form $\eta = d \star \omega$ are given by

$$\eta_{l_1 \ldots l_{n-p+1}} = \frac{1}{(n-p)!p!} \sum_{\substack{k, j_1 \ldots j_p \\ i_{p+1} \ldots i_n}}^{n} \delta^{l_1 \ldots l_{n-p+1}}_{k\, i_{p+1} \ldots i_n} \epsilon_{j_1 \ldots j_p i_{p+1} \ldots i_n} \frac{\partial}{\partial x^k} \left(\sqrt{|g|}\, \omega^{j_1 \ldots j_p} \right) \tag{B.2}$$

It remains to act with the $\star$ operator on η. This yields

$$
\begin{aligned}
\star\eta &= \frac{\sqrt{|g|}}{(n-p+1)!(p-1)!} \sum_{\substack{l_1\ldots l_{n-p+1}\\ m_1\ldots m_{p-1}}}^{n} \eta^{l_1\ldots l_{n-p+1}} \epsilon_{l_1\ldots l_{n-p+1}m_1\ldots m_{p-1}} dx^{m_1}\wedge\ldots dx^{m_{p-1}} \\
&= \frac{\sqrt{|g|}}{(n-p+1)!(p-1)!} \sum_{\substack{l_1\ldots l_{n-p+1}\\ m_1\ldots m_{p-1}\\ r_1\ldots r_{p-1}}}^{n} \eta_{l_1\ldots l_{n-p+1}} \epsilon^{l_1\ldots l_{n-p+1}r_1\ldots r_{p-1}} g_{r_1 m_1}\cdots g_{r_{p-1}m_{p-1}} \\
&\quad \times dx^{m_1}\wedge\ldots\wedge dx^{m_{p-1}} \\
&= \frac{\sqrt{|g|}}{g} \frac{1}{(n-p+1)!(p-1)!} \sum_{\substack{l_1\ldots l_{n-p+1}\\ m_1\ldots m_{p-1}\\ r_1\ldots r_{p-1}}}^{n} \eta_{l_1\ldots l_{n-p+1}} \epsilon_{l_1\ldots l_{n-p+1}r_1\ldots r_{p-1}} g_{r_1 m_1}\cdots g_{r_{p-1}m_{p-1}} \\
&\quad \times dx^{m_1}\wedge\ldots\wedge dx^{m_{p-1}}
\end{aligned}
$$

where in the last step we used Eq.(B.1). The only thing that remains is to insert the explicit form (B.2) into this expression. This gives

$$
\begin{aligned}
\star d\star\omega &= \frac{\mathrm{sign}(g)}{\sqrt{|g|}} \frac{1}{p!(n-p)!(n-p+1)!(p-1)!} \sum_{\substack{l_1\ldots l_{n-p+1}\\ m_1\ldots m_{p-1}, r_1\ldots r_{p-1}\\ k, j_1\ldots j_p, i_{p+1}\ldots i_n}}^{n} \delta^{l_1\ldots l_{n-p+1}}_{k\, i_{p+1}\ldots i_n} \\
&\quad \times \epsilon_{j_1\ldots j_p i_{p+1}\ldots i_n} \epsilon_{l_1\ldots l_{n-p+1}r_1\ldots r_{p-1}} \frac{\partial}{\partial x^k}\left(\sqrt{|g|}\,\omega^{j_1\ldots j_p}\right) g_{r_1 m_1}\cdots g_{r_{p-1}m_{p-1}} dx^{m_1}\wedge\ldots\wedge dx^{m_{p-1}} \\
&= \frac{\mathrm{sign}(g)}{\sqrt{|g|}} \frac{1}{p!(n-p)!(p-1)!} \sum_{\substack{m_1\ldots m_{p-1}, r_1\ldots r_{p-1}\\ k, j_1\ldots j_p, i_{p+1}\ldots i_n}}^{n} \frac{\partial}{\partial x^k}\left(\sqrt{|g|}\,\omega^{j_1\ldots j_p}\right) \\
&\quad \times \epsilon_{j_1\ldots j_p i_{p+1}\ldots i_n} \epsilon_{k\, i_{p+1}\ldots i_n r_1\ldots r_{p-1}} g_{r_1 m_1}\cdots g_{r_{p-1}m_{p-1}} dx^{m_1}\wedge\ldots\wedge dx^{m_{p-1}} \\
&= \frac{\mathrm{sign}(g)}{\sqrt{|g|}} \frac{(-1)^{(n-p)(p-1)}}{p!(n-p)!(p-1)!} \sum_{\substack{m_1\ldots m_{p-1}, r_1\ldots r_{p-1}\\ k, j_1\ldots j_p, i_{p+1}\ldots i_n}}^{n} \frac{\partial}{\partial x^k}\left(\sqrt{|g|}\,\omega^{j_1\ldots j_p}\right) \\
&\quad \times \epsilon_{j_1\ldots j_p i_{p+1}\ldots i_n} \epsilon_{k\, r_1\ldots r_{p-1} i_{p+1}\ldots i_n} g_{r_1 m_1}\cdots g_{r_{p-1}m_{p-1}} dx^{m_1}\wedge\ldots\wedge dx^{m_{p-1}}
\end{aligned}
\tag{B.3}
$$

where in the last step we shifted $n-p$ indices over $p-1$ positions. If we now use that

$$
\sum_{i_1\ldots i_{p+1}}^{n} \epsilon_{j_1\ldots j_p i_{p+1}\ldots i_n} \epsilon_{k\, r_1\ldots r_{p-1} i_{p+1}\ldots i_n} = (n-p)!\,\delta^{k\, r_1\ldots r_{p-1}}_{j_1\ldots j_p}
$$

we find

$$\star\, d \star \omega = \frac{\mathrm{sign}(g)}{\sqrt{|g|}} \frac{(-1)^{(n-p)(p-1)}}{p!(p-1)!} \sum_{\substack{m_1 \dots m_{p-1}, r_1 \dots r_{p-1} \\ k, j_1 \dots j_p}}^{n} \delta^{k\, r_1 \dots r_{p-1}}_{j_1 \dots j_p} \frac{\partial}{\partial x^k} \left(\sqrt{|g|}\, \omega^{j_1 \dots j_p} \right)$$

$$\times\, g_{r_1 m_1} \dots g_{r_{p-1} m_{p-1}} dx^{m_1} \wedge \dots \wedge dx^{m_{p-1}}$$

$$= \frac{\mathrm{sign}(g)}{\sqrt{|g|}} \frac{(-1)^{(n-p)(p-1)}}{(p-1)!} \sum_{\substack{m_1 \dots m_{p-1} \\ k, r_1 \dots r_{p-1}}}^{n} \frac{\partial}{\partial x^k} \left(\sqrt{|g|}\, \omega^{k\, r_1 \dots r_{p-1}} \right)$$

$$\times\, g_{r_1 m_1} \dots g_{r_{p-1} m_{p-1}} dx^{m_1} \wedge \dots \wedge dx^{m_{p-1}}$$

$$= \frac{1}{(p-1)!} \sum_{m_1 \dots m_{p-1}}^{n} (\star\, d \star \omega)_{m_1 \dots m_{p-1}} dx^{m_1} \wedge \dots \wedge dx^{m_{p-1}} \tag{B.4}$$

where we defined

$$(\star\, d \star \omega)_{m_1 \dots m_{p-1}} = \frac{\mathrm{sign}(g)}{\sqrt{|g|}} (-1)^{n(p-1)} \sum_{k, r_1 \dots r_{p-1}}^{n} \frac{\partial}{\partial x^k} \left(\sqrt{|g|}\, \omega^{k\, r_1 \dots r_{p-1}} \right) g_{r_1 m_1} \dots g_{r_{p-1} m_{p-1}}$$

where we used that $(-1)^{p(p-1)} = 1$. This expression is equivalent to Eq.(7.61).

Bibliography

[1] A. Einstein, "Zur Elektrodynamik bewegter Körper", Annalen der Physik **17**, (1905)

[2] A. Einstein, "Über die spezielle und die allgemeine Relativitätstheorie", (1916)

[3] G. Rizzi, M. L. Ruggiero, and A. Serafini, "Synchronization Gauges and the Principles of Special Relativity", Foundations of Physics, **34**, 1835, (2004)

[4] M. B. Cranor, E. M. Heider and R. H. Price, "A circular twin paradox", Am. J. Phys. **68**, 1016 (2000)

[5] K. Kassner, "Spatial geometry of the rotating disc and its non-rotating counterpart", Am. J. Phys. **80**, 772 (2012)

[6] L. Landau and E. Lifshitz, "The classical theory of fields"

[7] W. K. H. Panofsky and M. Phillips, "Classical Electricity and Magnetism", Addison-Wesley, (1977)

[8] H. A. Lorentz, "The Theory of Electrons", (Teubner, Leipzig, 1909)

[9] A. D. Yaghjian, "Relavistic Dynamics of a Charged Sphere: Updating the Lorentz-Abraham Model", Springer (2006)

[10] E. Poisson, "An introduction to the Lorentz-Dirac equation", arXiv:gr-qc/9912045 (1999)

[11] C. Teitelboim, D. Villarroel, Ch. G. van Weert, "Classical Electrodynamics of Retarded Fields and Point Particles", Revista del Nuovo Cimento, **3**, 9 (1980)

[12] R. Rivera and D. Villaroel,"Exact solutions to the Mo-Papas and Landau-Lifshitz equations", Phys. Rev. E**66**, 046618 (2002)

[13] Y. Kravets, A. Noble and D. Jaroszynski, "Radiation effects on the interaction of an electron with an intense laser pulse", Phys. Rev. E**88**, 011201(R), (2013)

[14] A. Di Piazza, K. Z. Hatsagortsyan, and C. H. Keitel, "Quantum Radiation Reaction Effects in Multiphoton Compton Scattering", Phys. Rev. Lett. **105**, 220403 (2010)