RT1: WS 08/09

Robert Beig

July 5, 2010

Contents

1 Introduction **3**
- 1.1 Continuum mechanics 3
- 1.2 Interlude on symmetry and 'covariance' 7
- 1.3 A remark on notation 13
- 1.4 Continuation 14
- 1.5 Maxwell theory 16
- 1.6 The Galilei spacetime M_G 19

2 Minkowski spacetime: the kinematics of special relativity **21**
- 2.1 Causal and spacelike vectors 21
- 2.2 Simultaneity 23
- 2.3 The reverse triangle inequality and the twin 'paradox' 24
- 2.4 Time dilation and gamma factor 26
- 2.5 Proper time 27
- 2.6 Length contraction 28
- 2.7 *Light: wave fronts and rays 29
- 2.8 *Phase and frequency 31
- 2.9 Doppler effect 32
- 2.10 Aberration of light 33

3 Relativistic mechanics **35**
- 3.1 Four momentum 35
- 3.2 Relativistic billiards 38
- 3.3 Compton effect 39
- 3.4 Accelerated motion 40
- 3.5 The relativistic rocket 41

4 Relativistic field theory **43**
- 4.1 Relativistic electrodynamics 43
- 4.2 The four potential A_μ 45
- 4.3 The Lienard-Wiechert fields 47
- 4.4 The energy momentum tensor 48
- 4.5 Energy momentum tensor of dust 49

4.6 Energy momentum tensor of perfect fluid matter 51

5 Basics of general relativity **53**
5.1 The equivalence principle 53
5.2 Curved spacetimes 54
5.3 Again: equivalence principle 58
5.4 The concept of curvature 60
5.5 Curvature in terms of the metric 62
5.6 The meaning of curvature 63
5.7 Newtonian limit of the geodesic equation 64
5.8 The Einstein field equations 65
5.9 The Schwarzschild spacetime 68
5.10 Gravitational Redshift 71
5.11 Geodesics of Schwarzschild 71
5.12 Relativistic cosmology 77
5.12.1 Cosmological Principle 77
5.12.2 The cosmological red-shift 79
5.12.3 The dynamics of the universe 80
5.12.4 The spatial geometry 81

6 Appendices **83**
6.1 A: Covariant differentiation 83
6.2 B: Symmetries of Euclidean space 85

Chapter 1

Introduction

1.1 Continuum mechanics

SRT historically emerged through a clash between symmetry properties of mechanics on one hand and those of the field theory of electromagnetism on the other hand. In order to understand the nature of this clash, it is useful to have a clear picture, first of the way in which mechanics can describe wave phenomena, secondly how different symmetries (actually: the Galilei transformations) are implemented in the description of these phenomena.
We consider a model for an elastic rod, i.e. a 1-dimensional object able to perform motions in the longitudinal direction. This model is very idealized, in particular we take the rod to be infinitely long.
Let $X = f(t,x)$ be the position in the undeformed configuration of the particle at location x for time t. Thus f has the meaning of 'inverse deformation'. We assume that (inverse) deformations preserve orientation so that $\partial_x f > 0$. We postulate that the variable $f(t,x)$ is subject to the following field equation

$$Mf := -\left(\partial_t - \frac{\partial_t f}{\partial_x f}\,\partial_x\right)\frac{\partial_t f}{\partial_x f} + c_s^2\,\partial_x^2 f = 0\,. \tag{1.1}$$

The complicated-looking first term essentially forms the acceleration-part of the equation (more about this later), whereas the second term describes the restoring elastic force. The latter is multiplied by a constant of dimension (velocity)^2 [1].
The natural ground state is given by

$$f_0(t,x) = x\,, \tag{1.2}$$

which of course corresponds to the material being at rest in the undeformed configuration. Clearly $f = f_0$ solves Eq.(1.1). We call f_0 the ether solution.
Eq.(1.1) is invariant under Galilei transformations in a sense explained below.

[1] Since f has dimension [length], $\partial_x f$ is dimensionless.

First recall Galilei transformations in 1 spatial dimension are given by the three-parameter family

$$(t,x) \mapsto (\bar{t},\bar{x}) = (t-\tau, x-\xi-vt) \tag{1.3}$$

When $\tau = v = 0$, these are space translations, when $\xi = v = 0$ these are time translations and when $\tau = \xi = 0$ we have boosts, i.e. transitions between different inertial systems. For the latter observe that the spatial origin of the new coordinate system $(\bar{t},\bar{x})$, i.e. $\bar{x} = 0$, is given by the line $(t,x)(\lambda) = (\lambda, v\lambda)$ in $\mathbb{R}^2$. This corresponds to an observer moving uniformly at velocity v through the origin $(0,0) \in \mathbb{R}^2$. We assume f transforms as a scalar under coordinate transformations, i.e. $\bar{f}$ is defined by $\bar{f}(\bar{t},\bar{x}) = f(t,x)$. Recall what this means: when $(\bar{t},\bar{x})$ is related to (t,x) in some way (in our case: 'by a Galilei transformation'), then the value of $\bar{f}$ at $(\bar{t},\bar{x})$ and that of (t,x) should be the same. Note that the previous statement does NOT mean that the functional dependence of $\bar{f}$ on $(\bar{t},\bar{x})$ is the same as that of f on (t,x), see the end of the section on symmetry and covariance. If, however, it **is** true that $\bar{f}$ on $(\bar{t},\bar{x})$ is the same as that of f on (t,x), we say that the state f is invariant under the transformation.
The operator M in (1.1) is invariant under Galilei transformations in the following sense: when $(\bar{t},\bar{x})$ is related to (t,x) by a Galilei transformation, then

$$(Mf)(t,x) = (M\bar{f})|_{(\bar{t}(t,x),\bar{x}(t,x))} \,. \tag{1.4}$$

This is easy to check for space and time translations. For boosts note first the identities (omitting arguments)

$$\partial_t g = \partial_{\bar{t}}\bar{g} - v\,\partial_{\bar{x}}\bar{g}, \qquad \partial_x g = \partial_{\bar{x}}\bar{g} \tag{1.5}$$

which follow from differentiating both sides of $g(t,x) = \bar{g}(\bar{t},\bar{x})$ and using the chain rule on the r.h. side. This implies that $\partial_x^2 f = \partial_{\bar{x}}^2 \bar{f}$, i.e. the second term in (1.1) alone satisfies (1.4). For the first term in (1.1) this requires a slightly lengthier calculation. We have, by (1.5), that

$$\frac{\partial_t f}{\partial_x f} = \frac{\partial_{\bar{t}}\bar{f}}{\partial_{\bar{x}}\bar{f}} - v \tag{1.6}$$

and thus that

$$\partial_t g - \frac{\partial_t f}{\partial_x f}\,\partial_x g = \partial_{\bar{t}}\bar{g} - \frac{\partial_{\bar{t}}\bar{f}}{\partial_{\bar{x}}\bar{f}}\,\partial_{\bar{x}}\bar{g} \tag{1.7}$$

Combining (1.6) with (1.7), we see that the first term in (1.1) also satisfies (1.4).

Exercise 1: Check in detail that both terms in M defined by (1.1) are invariant under boosts in the sense that (1.4) is valid, whenever $\bar{t} = t$, $\bar{x} = x - vt$. e.o.e. (end of exercise)

Since a general Galilei transformation can be written as the composition of a time translation, a space translation and a boost, we are done. It follows from (1.4) that, when f is a solution of the theory, so is $\bar{f}$.
We now explain the sense in which the first term is an acceleration. Namely let $x = F(t, X)$, i.e. the map sending the point X in the undeformed configuration to its location at time t, i.e. $f(t, F(t, X)) = X$. Then we have the identities

$$u(t, x) := \partial_t F(t, X)|_{X=f(t,x)} = -\frac{\partial_t f(t, x)}{\partial_x f(t, x)} \tag{1.8}$$

and

$$b(t, x) := \partial_t^2 F(t, X)|_{X=f(t,x)} = -\left(\partial_t - \frac{\partial_t f(t, x)}{\partial_x f(t, x)}\,\partial_x\right)\frac{\partial_t f(t, x)}{\partial_x f(t, x)} \tag{1.9}$$

The proof of these relations follows from differentiating $f(t, F(t, X)) = X$ with respect to t and X and using the chain rule. The relations (1.8,1.9) also show that the acceleration term in (1.1) is nothing but (minus) the $\partial_t u + (u\nabla)u$ - term of Eulerian hydrodynamics.

Exercise 2: Check (1.9). e.o.e.

In this connection it may be instructive to recall the other fundamental law, namely the equation of continuity which in nonrelativistic continuum mechanics describes conservation of mass. In the present formulation this arises as follows. Let $\rho_0(X)$ be the mass density in the undeformed configuration. Then, by the change-of-variable-law for integrals, the actual mass density is given by $\rho(t, x) = \rho_0(f(t, x))\,\partial_x f(t, x)$. After a short calculation we find that the equation

$$\partial_t \rho + \partial_x(\rho u) = 0 \tag{1.10}$$

holds as an identity.

Exercise 3: Check (1.10).e.o.e.

Equation (1.1) permits a further symmetry, namely it is the case that the map N given by

$$Nf = f - a\,, \tag{1.11}$$

where a is a constant, sends solutions into solutions. We can also for example combine this with a translation in x, i.e. $\bar{t} = t, \bar{x} = x - \xi$, by defining

$$\bar{f}(\bar{t}, \bar{x}) = f(t, x) - \xi = f(\bar{t}, \bar{x} + \xi) - \xi\,, \tag{1.12}$$

and this is again a symmetry of the theory in the sense that Eq.(1.4) is valid.
Let us apply these concepts to the ether solution. Under a time translation we

clearly have that $\bar{f}_0(\bar{t},\bar{x}) = f_0(t,x) = x = \bar{x}$. Thus time translations, in addition to being symmetries of the theory, leave the ether solution invariant. Similarly the ether solution f_0 is invariant under the combined translational symmetry defined in (1.12), since $\bar{f}_0(\bar{t},\bar{x}) = f_0(\bar{t},\bar{x}+\xi) - \xi = \bar{x}+\xi-\xi = \bar{x}$.
Finally turning to boosts, we have that

$$\bar{f}_0(\bar{t},\bar{x}) = f_0(t,x) = x = \bar{x} + v\bar{t} \tag{1.13}$$

Thus boosts, which are symmetries of the theory, do NOT leave f_0 invariant[2]. Thus we find a whole family of new solutions of Eq.(1.1), namely

$$f_0(t,x;v) = x + vt\,. \tag{1.14}$$

These results are physically very plausible. The ether solution looks the same everywhere and at all times. But it DOES matter, if I observe it in its rest system or in a system which is boosted relative to its rest system. (One might describe this even more pictorially by saying "I feel the ether wind".) Alternatively, in the active interpretation, it makes a difference if the ether is at rest with respect to me or moving at velocity v.
Next we look at small excitations of the ground state afforded by the ether solution. This is done as follows. Suppose we have have a family $f_\lambda(t,x)$ of solutions to (1.1), so that $f_0(t,x) = x$. Define

$$\delta f(t,x) = \frac{d}{d\lambda} f_\lambda(t,x)|_{\lambda=0} \tag{1.15}$$

Differentating $Mf_\lambda = 0$ w.r. to λ and setting λ to zero, we find that

$$\frac{d}{d\lambda} M(f_\lambda)|_{\lambda=0} = -\partial_t^2 \delta f + c_s^2\, \partial_x^2 \delta f = 0\,, \tag{1.16}$$

i.e. the linear perturbation δf obeys the wave equation with c_s playing the role of phase velocity.

Exercise 4: Prove (1.16).e.o.e.

For purposes which will become clear later, we write (1.16) as follows. We take Greek indices $\mu, \nu, ...$ with values $0, 1$ and define $x^\mu = (t,x)$. Furthermore we define a symmetric matrix of the form

$$\eta^{\mu\nu} = \begin{pmatrix} -\frac{1}{c_s^2} & 0 \\ 0 & 1 \end{pmatrix}$$

[2] The trick similar to that used for translations of adding a term, in the definition of $\bar{f}$, to make up for the last term in (1.13) does not work: adding a term proportional to t is not a symmetry of the theory.

Then (1.16) can be rewritten as

$$\Box\,\delta f := \sum_{\mu,\,\nu} \eta^{\mu\nu}\partial_\mu\partial_\nu\,\delta f = 0 \tag{1.17}$$

The operator $\Box$ is called the d'Alembert equation and should be seen in formal analogy to the familiar Laplace operator. Equation (1.16) is invariant under space and time translations, but NOT under boosts. This is not unreasonable: when we linearize a theory with a certain symmetry at a solution which is not invariant under this symmetry, the linearized theory has no reason to inherit this symmetry. Another aspect of this phenomenon is to linearize our theory at a boosted ether solution, namely $f_0(t, x; v)$. The result is the equation

$$\sum_{\mu,\,\nu} g^{\mu\nu}\partial_\mu\partial_\nu\,\delta f = 0\,, \tag{1.18}$$

where

$$g^{\mu\nu} = \begin{pmatrix} -\frac{1}{c_s^2} & \frac{v}{c_s^2} \\ \frac{v}{c_s^2} & 1 - \frac{v^2}{c_s^2}\,. \end{pmatrix}$$

Exercise 5: Check the above claim. The object $g^{\mu\nu}$ is sometimes called the acoustic metric.e.o.e.

In accordance with the law of velocity addition of Galilean physics, it decribes - provided that $|v| < c_s$ - right-traveling waves with phase velocity $c_s + v$ and left-traveling waves with phase velocity $c_s - v$. To check this we make the ansatz

$$\delta f \sim e^{i(\omega t + kx)} \tag{1.19}$$

with the result that $\omega = -kv \pm |k|c$.
We have in this section obtained the standard wave equation which, as is well-known and as we will show shortly, is a "relativistic" equation, from a nonrelativistic theory. And indeed: if we transform, within this non-relativistic theory, the wave equation between different inertial systems, it changes its form. Another point worth mentioning is this: the appearance of the standard wave equation within the above theory is in a sense an artefact of restricting oneself to one dimension. Had we started from a full three dimensional theory of elasticity we would have obtained a more complicated equation describing both longitudinal and transversal waves, each traveling with their own velocity.

1.2 Interlude on symmetry and 'covariance'

We now pause for some more elaboration of the concept of symmetry vis a vis the concept of 'covariance'. This will be a significant detour from our main path, but will turn out to be a good investment - at least once we come to GRT. It has to be said right away that the concept of 'covariance' used here follows unfortunate usage in the physics literature. It is not meant to be in the sense of "something transforms covariantly" as opposed to "contravariantly" or in a mixed fashion. Rather the term covariance in the present context includes all these possibilities. We already had the concept of symmetry of a theory, e.g. "the theory given by Eq.(1.1) is Galilei invariant" and that of the symmetry of a solution to such a theory, e.g. "the ether solution is invariant under translations in time and space". Now by "covariance" one means the phenomenon that certain operations (such as for example appearing in the field equations of some theory) keep their form under arbitrary coordinate transformations. For which theories, if any, does this hold? The answer at this point is that the question does not make sense. Rather the question should have been "can we give transformations rules for the quantities of a theory so that the expressions in which they appear are covariant?", and then it is plausible that this can always be done, and presumably in many ways! In fact, at least for the fundamental theories of physics there is always a way which in everybody's view is natural, and this basically - in modern language - consists of making explicit the geometrical structures entering the theory and finding out how these structures transform under arbitrary coordinate transformations. Let us give a rough sketch of this procedure in terms of some examples which should in part be familiar. Suppose we have a vector field $\mathbf{v}(\mathbf{x})$ and a scalar field $\sigma(\mathbf{x})$ with $\mathbf{x} \in \mathbb{R}^3$. Consider the expression

$$(\mathbf{v}\nabla)\,\sigma = (v^1\partial_1 + v^2\partial_2 + v^3\partial_2)\,\sigma \tag{1.20}$$

where we use the shorthand ∂_1 for partial derivative w.r. to x^1, a.s.o. This expression has the interpetation of "rate of change of the scalar field σ along the flow defined by the vector field $\mathbf{v}$". In other words: let $\mathbf{y}(t)$ be an integral curve of $\mathbf{v}$, i.e. $\mathbf{v}(\mathbf{x}(t)) = \dot{\mathbf{x}}(t)$. Now let $\Sigma(t)$ be the restriction of σ to this curve. Then, by the chain rule, expression (1.20) is just the t-derivative of Σ. Now we ask what happens if we allow arbitrary coordinate transformations. We already know how vector fields transform under coordinate transformations $\mathbf{x} \mapsto \bar{\mathbf{x}} = \mathbf{f}(\mathbf{x})$ or, in index notation,

$$\bar{x}^i = f^i(x^j)\,, \tag{1.21}$$

where i, j run from 1 to 3. Namely the rule is that the components $\bar{v}^i$ of the vector field in the new coordinates are given by ("contravariant behaviour")

$$\bar{v}^i(\mathbf{f}(\mathbf{x})) = \sum_{j=1}^{3}(\partial_j f^i)(\mathbf{x})\,v^j(\mathbf{x}) \tag{1.22}$$

Next recall that σ transforms like scalar, i.e.

$$\bar{\sigma}(\mathbf{f}(\mathbf{x})) = \sigma(\mathbf{x}) \tag{1.23}$$

Consequently

$$\sum_{j=1}^{3} \bar{v}^j(\mathbf{f}(\mathbf{x}))(\partial_j\bar{\sigma})(\mathbf{f}(\mathbf{x})) = \sum_{i=1}^{3} v^i(\mathbf{x})(\partial_i\sigma)(\mathbf{x}) \tag{1.24}$$

Thus the object σ defined by expression (1.20) transforms again like a scalar field. So, in this special example the answer to our question "what do we have to do to render the expression covariant?" is "nothing: this is already taken care of by assuming σ is a scalar and v^i a vector."

Exercise 6: For the operator O, sending the scalar field σ into $\sum_i v^i\partial_i\sigma$ we can again ask the "symmetry-question": what does it take, for a transformation $x^i \mapsto \bar{x}^i = f^i(x)$, to preserve the operator O in the sense of (1.4), i.e. $(O\sigma)(x) = (O\bar{\sigma})(f(x))$? Show the answer is that $v^i(x)$ satisfies the relation $v^i(f(x)) = \sum_j(\partial_j f^i)(x)v^j(x)$.
Solution: By the relation $\sigma(x) = \bar{\sigma}(\bar{x})$, we have that

$$(O\sigma)(x) = (v^i(x)\partial_i\sigma)(x) = v^i(x)\partial_i\bar{\sigma}(\bar{x}(x)) = v^i(x)\frac{\partial\bar{x}^j}{\partial x^i}(x)\bar{\partial}_j\bar{\sigma}(\bar{x}(x)) \tag{1.25}$$

On the other hand we have that

$$O\bar{\sigma}(\bar{x}) = v^i(\bar{x})\bar{\partial}_i\bar{\sigma}(\bar{x}) \tag{1.26}$$

Now the claim results from comparing coefficients.e.o.e.

We can also look at expression (1.20) 'without v^i': the quantity $\partial_i\sigma$, i.e. the gradient of the scalar σ, has a specific transformation behaviour, namely $\partial_i\sigma = \frac{\partial f^j}{\partial x^i}\bar{\partial}_j\bar{\sigma}$. We next point out (recall) that there are many fields whose transformation property is that of the gradient of a scalar without necessarily being gradients of scalars. They are called covector fields or 1-forms and are denoted with downstairs indices, e.g. $\omega_i(\mathbf{x})$. So their behaviour is ("covariant behaviour": note the inconsistency with the term "covariance" used in the section heading!)

$$\omega_i(\mathbf{x}) = \sum_{j=1}^{3}(\partial_i f^j)(\mathbf{x})\,\bar{\omega}_j(\mathbf{f}(\mathbf{x})) \tag{1.27}$$

We will from now on apply the "Einstein summation convention" to the effect that summation over pairs of indices, one up one down, is understood without the need for the summation sign. Then, together with some sloppiness concerning arguments of functions, we may write (1.22) as

$$\bar{v}^i = \frac{\partial\bar{x}^i}{\partial x^j}\,v^j \tag{1.28}$$

and (1.27) as

$$\omega_i = \frac{\partial\bar{x}^j}{\partial x^i}\,\bar{\omega}_j\,, \tag{1.29}$$

what is the same as

$$\bar{\omega}_i = \frac{\partial x^j}{\partial \bar{x}^i}\,\omega_j \tag{1.30}$$

It is clear that these two spaces, vector fields and covector fields, are linear spaces, where the neutral element and addition are defined in the obvious sense, scalars are defined as scalar fields(!), and multiplication by scalars again in the obvious sense.

Exercise 7: Show that every covector is a linear combination (with coefficients scalar fields: not scalars) of gradients of scalar fields. Hint: For the covector field with components $\omega_i(x) = (\omega_1(x), 0, 0, .., 0)$ there holds $\omega_i(x) = \omega_1(x)\partial_i\sigma(x)$ for a suitable scalar field σ. e.o.e.

It also follows from the above that these two spaces are not independent of each other. There is what mathematicians call a "natural pairing" between them. Namely the expression $v^i\omega_i = \omega_i v^i$ defines, for given ω_i, a linear map, sending each vector field into a scalar field. It is also not very hard to see that every such linear map can be gotten from a covector field ω_i. One can in addition read $v^i\omega_i$ as defining a linear map from covector fields to scalar fields and again check that any such map can be gotten from a vector field. In the language of linear algebra we have thus seen that "covector fields form the vector space dual to vector fields" and, furthermore, "the dual space of covector fields is the same as the space of vector fields". As a note of caution we add that all these spaces have infinite dimension, but, when attention is confined to some fixed point $\mathbf{x}$, each is three dimensional. Nothing prevents us from doing the above in n rather than three dimensions. Then of course the space of vectors and covectors at a point becomes n-dimensional.
We are now in a position to define more general tensors - in essentially two ways. Take tensor fields of valence (2,0) for example: these are given by all linear combinations of pointwise products ('tensor products") of components of vector fields, i.e.

$$t^{ij} = \sigma v^i w^j + ... + ... \tag{1.31}$$

It is thus clear that the space of such tensors[3] at some given point has dimension n^2. Knowing that "vectors are dual to covectors", we can, alternatively and more abstractly, define these tensors as bilinear maps from the cartesian product of the space of vector fields with itself to the space of scalar fields.
Along similar lines we can define tensor fields of arbitrary valence (p, q). Clearly vector fields are tensor fields of valence $(1, 0)$ and covector fields are tensor fields of valence $(0, 1)$.

[3]When no confusion results, we will often use the words "tensor field" and "tensor" synonymously.

Exercise 8: Show that in a suitable sense tensor fields of valence (p, q) behave contravariantly w.r. to upper indices and covariantly w.r. to lower indices.e.o.e.
Exercise 9: Define a tensor field $t^i{}_j$ of valence (1,1) by $t^i{}_j = \delta^i{}_j$ in some coordinate system, where $\delta^i{}_j$ is the Kronecker delta. Show that this particular tensor field is invariant in the sense of having the same value in all coordinate systems.e.o.e.
Exercise 10: Show that there is an obvious concept of tensor product yielding, for a pair of tensors of valence (p, q) and (p', q') one of valence $(p + p', q + q')$. Show that this tensor product is not commutative, despite that for example $u^i v^j = v^j u^i$.e.o.e.
Exercise 11: We may view the pairing between valence-$(1, 0)$ and valence-$(0, 1)$ tensors as a process of "contraction" of the $(1, 1)$ tensor given by the tensor product between these respective tensors. Show there is a process of contraction generalizing this which, for tensor fields of valence $(p+1, q+1)$, gives well defined tensor fields of valence (p, q). As an additional check verify this is consistent with the transformation rules for $(p + 1, q + 1)$ and $(p.q)$ tensors.e.o.e.

This ends our crash course on tensor fields. As for our next example imagine we have some time independent potential given by a scalar field $U(\mathbf{x})$ and look at the Laplace operator defined as

$$\Delta U = (\partial_1^2 + \partial_2^2 + \partial_3^2)U = \delta^{ij}\partial_i\partial_j U\,, \tag{1.32}$$

where δ^{ij} is again the Kronecker delta. What should be the coordinate-independent meaning of this operation? The "right " answer turns out to be that suggested by the index structure of expression (1.32). Namely one first interprets δ^{ij} as a (2,0) tensor, second interprets $\partial_i\partial_j$ as a (0,2) tensor, and then the whole expression will simply be a double contraction between the two. We start with δ^{ij}. In contrast to the Kronecker delta viewed as a (1,1) tensor, the Kronecker delta viewed as a (2,0) tensor changes its form under general coordinate transformations.
Example: Take the transformation, the inverse of which is given by

$$x^1 = r\sin\theta\cos\phi,\ x^2 = r\sin\theta\sin\phi,\ x^3 = r\cos\theta \tag{1.33}$$

Calling, for consistency, $\delta^{ij} = h^{ij}$ and the same object in the new coordinates $\bar{h}^{ij}$, where $\bar{x}^1 = r, \bar{x}^2 = \theta, \bar{x}^3 = \phi$, we find after some calculations that

$$\bar{h}^{ij} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{r^2} & 0 \\ 0 & 0 & \frac{1}{r^2\sin^2\theta} \end{pmatrix}$$

Since h^{ij} is invertible, there exists a unique tensor h_{ij}, also positive definite, such that

$$h^{ij}h_{jk} = \delta^i{}_k \tag{1.34}$$

in the original coordinate system this is again the Kronecker delta, whereas in polar coordinates it reads

$$\bar{h}_{ij} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & r^2 & 0 \\ 0 & 0 & r^2 \sin^2 \theta \end{pmatrix}$$

which is of course nothing but the components of the Euclidean line element in spherical coordinates.

Exercise 12: Verify the above covariant form of the Euclidean metric in spherical coordinates.e.o.e.

It remains to see what becomes of $\partial_i \partial_j U$ in general coordinates. While the gradient of a scalar transforms as a covector, as we have seen, the partial derivative $\partial_i \omega_j$ of a covector ω_i does not transform like a (0,2)-tensor. We have to seek an expression $D_i \omega_j$, which does transform correctly and reduces to $\partial_i \omega_j$ in the original Euclidean coordinates. As we show in Appendix A, the answer is given by the following formula

$$D_i \omega_j = \partial_i \omega_j - \Gamma^k_{ij} \omega_k \,, \tag{1.35}$$

where

$$\Gamma^i_{jk} = \frac{1}{2} h^{il} [\partial_j h_{kl} + \partial_k h_{jl} - \partial_l h_{jk}] \tag{1.36}$$

In the original coordinates the (contravariant) metric had constant components, so the "Christoffel symbol" Γ^i_{jk} is zero.
Now, then, that we have managed to write our expressions - or at least some representative ones - in general coordinates: what does it mean for a transformation to be a symmetry of this expression? It means that the functional dependence of the object in question on coordinates does not change under the change of coordinates. So in the case of a scalar field U this means that $\bar{U}$, which is defined by $\bar{U}(\bar{x}) = U(x)$ satisfies $\bar{U}(\bar{x}) = U(\bar{x})$, what is the same as $U(\bar{x}) = U(x)$. In the case of a (2,0) tensor field $t^{ij}(x)$ the condition on $\bar{x}^i(x^j)$ to be a symmetry becomes

$$\frac{\partial \bar{x}^i}{\partial x^k}(x) \frac{\partial \bar{x}^j}{\partial x^l}(x)\, t^{kl}(x) = t^{ij}(\bar{x}) \tag{1.37}$$

Let us apply this concept to the Euclidean $h^{ij} = \delta^{ij}$. It then follows from Appendix B, that $\bar{x}^i$ has to be of the form

$$\bar{x}^i = R^i{}_j x^j + d^i \,, \tag{1.38}$$

where d^i and $R^i{}_j$ are constants, the latter subject to

$$R^i{}_k R^j{}_l \, \delta^{kl} = \delta^{ij} \,. \tag{1.39}$$

Thus we find the Euclidean group, i.e. "translations plus rotations", as expected. Let us come back to the Laplace operator. It follows from the above that symmetries of $h^{ij} = \delta^{ij}$ are also symmetries of the Laplace operator. In other words, for every symmetry $f^i(x)$ it holds that $\Delta(U \circ f)(x) = (\Delta U)(f(x))$, and this can of course be checked explicitly from (1.38). We state without proof that every symmetry of the Laplace operator has to be a symmetry of h^{ij}.

A further exercise could be to understand how the elastic operator in (1.1) can naturally be written in arbitrary coordinates (t, x). This would be a bit complicated, so we refrain from doing this here.

1.3 A remark on notation

We have in the previous section dealt with covariant and contravariant objects of different types. There is a convenient notation for such objects which we will occasionally use, namely for a vector field v with components v^i in some coordinate system x^i one sometimes writes

$$v = v^i(x)\frac{\partial}{\partial x^i} = v^i(x)\partial_i\,, \tag{1.40}$$

where of course the summation convention is used. Thus a vector is seen as a first-order partial differential operator. There are two advantages in employing this notation: first the behavior of the components v^i under change of coordinates $x^i \to f^i(x^j)$ is automatic (prove this!). Second, the above notation is a nice bookkeeping device for specifying the components of the vector field v in some coordinate system. For covectors ω one sometimes writes

$$\omega = \omega_i(x)\,dx^i \tag{1.41}$$

For second-rank, symmetric covariant tensors, say t, one writes

$$t = t_{ij}(x)\,dx^i dx^j \tag{1.42}$$

where the object $dx^i dx^j$ is identified with $dx^j dx^i$. For example the covariant tensor with components

$$t_{ij} = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$$

in the coordinates (x, y) becomes

$$t = a\,dx^2 + 2b\,dxdy + c\,dy^2 \tag{1.43}$$

In practice this last notation is mainly used for metrics.

1.4 Continuation

Suppose next that we are looking at equation (1.16) not as some approximation to a (nonlinear) theory, but as a theory in its own right. If so, we are obliged to take the symmetries of this theory seriously. We again assume δf transforms like a scalar. As mentioned before, Eq.(1.16) is then invariant under space and time translations. In fact, in complete analogy to the previous section we can say, that every symmetry of the tensor $\eta^{\mu\nu}$ is a symmetry of the d'Alembert operator. The latter, as follows from the general result in Appendix B, consist of space and time translations plus linear transformations, i.e. transformations $(x^\mu) \mapsto (\bar{x}^\mu)$ of the form

$$\bar{x}^\mu = L^\mu{}_\nu \, x^\nu \tag{1.44}$$

Then (1.16), resp. (1.17) is invariant under this transformation, iff

$$\eta^{\mu\nu} = \eta^{\rho\sigma} L^\mu{}_\rho L^\nu{}_\sigma \tag{1.45}$$

Setting

$$L^\mu{}_\nu = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

there results

$$\frac{A^2}{c_s^2} - B^2 = 1, \quad \frac{C^2}{c_s^2} - D^2 = -1, \quad AC = BDc_s^2 \tag{1.46}$$

These are 3 equations for 4 unknowns. We are interested in 1-parameter family of solutions containing the identity. Here is the unique answer: there is a number $\xi \in \mathbb{R}$ such that

$$\frac{A}{c_s} = D = \cosh\,\xi, \quad B = \frac{C}{c_s} = -\sinh\,\xi \tag{1.47}$$

Now the new spatial origin $\bar{x}^1 = \bar{x} = 0$, i.e. $Ct + Dx = 0$, takes the form $t\,c_s \sinh\,\xi - x\,\cosh\,\xi = 0$ and thus, viewed from the (t, x) - system, moves at velocity $v = c_s \tanh\,\xi$ with $v/c_s \in (-1, 1)$ and the transformation (1.44) take the form

$$\bar{t} = \gamma(v)\left(t - \frac{v}{c_s^2}\,x\right), \quad \bar{x} = \gamma(v)(x - vt)\,, \tag{1.48}$$

where $\gamma(v)$ is given by

$$\gamma(v) = \frac{1}{\sqrt{1 - \frac{v^2}{c_s^2}}}\,. \tag{1.49}$$

Exercise 13: Check that

$$\eta^{\mu\nu}\bar{\partial}_\mu\bar{\partial}_\nu\bar{\phi} = \eta^{\mu\nu}\partial_\mu\partial_\nu\phi \tag{1.50}$$

where $\bar{\phi}(\bar{t}, \bar{x}) = \phi(t, x)$ and $(\bar{t}, \bar{x})$ is related to (t, x) by a Lorentz boost.e.o.e.

When we formally let $c_s \to \infty$, these transformations become Galilei boosts. In the process we see that they have the property of leaving the degenerate object

$$h^{\mu\nu} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

invariant. The most conspicuous feature of (1.48) is that both x AND t get transformed. Thus simultaneity of two 'events', e.g. of $(0, x)$ and $(0, x')$, is not preserved under these transformations.
Our considerations so far pretended the world has only one spatial dimension. But this was only done for ease of presentation. Similar things hold for 3 spatial dimensions. Then we would, for the small excitations $\phi(t, \vec{x})$, obtain the equation

$$\Box\phi := \left(-\frac{1}{c_s^2}\,\partial_t^2 + \Delta\right)\phi = 0\,, \tag{1.51}$$

In the following we will view the equation (1.51) as a fundamental equation in its own right, whose symmetries have to be taken seriously. We do this despite the fact that there, at least in macroscopic physics, no fundamental scalar field known. Recall that in theories of continuum mechanics, where such fields do arise, (1.51) appears as the equation governing linear perturbations off some ("ether") background field.
We now set $c_s = c$ and again write

$$\Box = \eta^{\mu\nu}\partial_\mu\partial_\nu\,, \tag{1.52}$$

where

$$\eta^{\mu\nu} = \begin{pmatrix} -\frac{1}{c^2} & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

In complete analogy we have that the operator sending the scalar field ϕ into the scalar field $\Box\phi$ has as symmetries all transformations of the form

$$\bar{x}^\mu = L^\mu{}_\nu x^\nu + d^\mu\,, \tag{1.53}$$

where d^μ and $L^\mu{}_\nu$ are constants, the latter subject to the condition

$$L^\mu{}_\rho L^\nu{}_\sigma \eta^{\rho\sigma} = \eta^{\mu\nu} \tag{1.54}$$

The set of these transformation, which form a group, is called the Poincaré group, the subgroup with $c^\mu = 0$ the Lorentz group. Here is an exhaustive list of elements of the Lorentz group. First we have

$$\bar{t} = t, \quad \bar{x}^i = R^i{}_j x^j\,, \tag{1.55}$$

where $i, j = 1, 2, 3$ and $R^i{}_j$ is in the 3 - parameter group of spatial rotations, i.e. satisfying

$$R^i{}_k R^j{}_l \delta^{kl} = \delta^{ij} \tag{1.56}$$

Another 1 - parameter subgroup is, as we know, given by Lorentz boosts in the x^1 direction

$$\bar{t} = \gamma(v^1)\left(t - \frac{v^1}{c^2}\,x^1\right), \;\; \bar{x}^1 = \gamma(v^1)(x^1 - v^1 t), \;\; \bar{x}^2 = x^2, \;\; \bar{x}^3 = x^3 \tag{1.57}$$

where $\gamma(v^1) = (1 - (\frac{v^1}{c})^2)^{-\frac{1}{2}}$. There are of course Lorentz boosts in all spatial directions. They can be either guessed or obtained by composing the above boost with suitable rotations. The result is

$$\bar{t} = \gamma(|\mathbf{v}|)\left(t - \frac{\mathbf{v}\mathbf{x}}{c^2}\right), \quad \bar{\mathbf{x}} = \mathbf{x} + \frac{\gamma(|\mathbf{v}|) - 1}{|\mathbf{v}|^2}\mathbf{v}(\mathbf{v}\mathbf{x}) - \gamma(|\mathbf{v}|)\mathbf{v}t \tag{1.58}$$

Note that the term $\frac{\gamma(|\mathbf{v}|)-1}{|\mathbf{v}|^2}$ extends to a smooth function of $|\mathbf{v}|^2$ also at zero velocity. Interestingly, the above transformations do not form a group. Namely, composing two boosts in different spatial directions we obtain a Lorentz transformation not of the form (1.58), but containing an additional rotation. This fact is important in connection with the phenomenon of Larmor precession.

Exercise 14: Check that (1.58) are Lorentz transformations, i.e. satisfy (1.54). e.o.e.
Exercise 15: Check that $L^\mu{}_\nu(\vec{v})$ defined by (1.58) satisfies $L^\mu{}_\nu(\vec{v})L^\nu{}_\rho(-\vec{v}) = \delta^\mu{}_\rho$. e.o.e.

1.5 Maxwell theory

Electromagnetism deals with two fields, $\mathbf{E}(t, \mathbf{x})$ and $\mathbf{B}(t, \mathbf{x})$ governed by the following system of linear partial differential equations:

$$\nabla\mathbf{E} = 4\pi\rho \qquad -\frac{1}{c^2}\,\partial_t\,\mathbf{E} + \nabla\times\mathbf{B} = 4\pi\mathbf{j} \tag{1.59}$$

and

$$\nabla\mathbf{B} = 0 \qquad -\partial_t\,\mathbf{B} - \nabla\times\mathbf{E} = 0 \tag{1.60}$$

Here ρ and $\mathbf{j}$ are respectively the densities of charge and current, the constant c is the speed of light in vacuum. These are the equations governing electromagnetic fields in vacuum and also their interaction with given charge and current distributions. Despite the fact that no traces of an ether were ever found experimentally (the strongest argument against its existence being the null result of the famous Michelson experiment) all leading researchers until Einstein believed in its existence. Maxwell himself had made efforts, without success, to find a mechanical interpretation for his own equations. If, then, we agree that the Maxwell theory is "fundamental", we are forced to take its invariances, if any are found, seriously. It turns out that the set of equations (1.59,1.60) are not Galilei invariant, but, in a sense to be explained, Lorentz invariant. Some evidence comes from the following consideration, for which we set the sources to be zero for simplicity. Let us take a time derivative of the second of (1.59) and use the second of (1.60) to eliminate $\partial_t \mathbf{B}$. Using a standard identity from vector analysis and the first of (1.60), we find that

$$\Box\, \mathbf{E} = 0\,, \tag{1.61}$$

and the same equation for the magnetic field.
We now explain the sense in which the operations occuring in the Maxwell equations are Lorentz invariant. The facts below were essentially known to Lorentz, and to some extent already to Larmor, but they were unable to draw the right conclusions. View the operations given by

$$m^\mu = \begin{pmatrix} m^0 \\ \mathbf{m} \end{pmatrix} = \begin{pmatrix} \nabla \mathbf{E} \\ -\partial_t \mathbf{E} + c^2 \nabla \times \mathbf{B} \end{pmatrix} \quad \text{and} \quad n^\mu = \begin{pmatrix} \nabla \mathbf{B} \\ -\partial_t \mathbf{B} - \nabla \times \mathbf{E} \end{pmatrix}$$

as defining maps sending pairs of three dimensional vector fields into 4-vectors on spacetime. Now consider Lorentz boosts given by (1.57) and for the electric and magnetic field take

$$\bar{\mathbf{E}} = \gamma(\mathbf{E} + \mathbf{v} \times \mathbf{B}) + \frac{1-\gamma}{|\mathbf{v}|^2}\mathbf{v}(\mathbf{v}\mathbf{E}) \qquad \bar{\mathbf{B}} = \gamma(\mathbf{B} - \frac{1}{c^2}\mathbf{v} \times \mathbf{E}) + \frac{1-\gamma}{|\mathbf{v}|^2}\mathbf{v}(\mathbf{v}\mathbf{B}) \tag{1.62}$$

We can check that these relations can be inverted in terms of $(\mathbf{E}, \mathbf{B})$ resulting in the same expression, but with the sign of $\mathbf{v}$ reversed, namely

$$\mathbf{E} = \gamma(\bar{\mathbf{E}} - \mathbf{v} \times \bar{\mathbf{B}}) + \frac{1-\gamma}{|\mathbf{v}|^2}\mathbf{v}(\mathbf{v}\bar{\mathbf{E}}) \qquad \mathbf{B} = \gamma(\bar{\mathbf{B}} + \frac{1}{c^2}\mathbf{v} \times \bar{\mathbf{E}}) + \frac{1-\gamma}{|\mathbf{v}|^2}\mathbf{v}(\mathbf{v}\bar{\mathbf{B}}) \tag{1.63}$$

From (1.58) we find that

$$\partial_t = \gamma\,(\partial_{\bar{t}} - \mathbf{v}\bar{\nabla}) \qquad \nabla = -\frac{\gamma}{c^2}\mathbf{v}\,\partial_{\bar{t}} + \bar{\nabla} + \frac{\gamma - 1}{|\mathbf{v}^2}\mathbf{v}(\mathbf{v}\bar{\nabla}) \tag{1.64}$$

Using (1.63) and (1.64) we find

$$\nabla \mathbf{E} = \gamma \bar{\nabla} \bar{\mathbf{E}} + \gamma \mathbf{v}[\bar{\nabla} \times \bar{\mathbf{B}} - \frac{1}{c^2}\partial_{\bar{t}} \bar{\mathbf{E}}] \tag{1.65}$$

But the r.h.side of (1.65) is nothing but $\gamma(\bar{m}^0 + \frac{1}{c^2}\mathbf{v}\bar{\mathbf{m}})$, and similarly a lengthy calculation would find that

$$\mathbf{m} = \gamma \mathbf{v} \bar{m}^0 + \frac{\gamma - 1}{|\mathbf{v}|^2}\mathbf{v}(\mathbf{v}\bar{\mathbf{m}}) + \bar{\mathbf{m}} \tag{1.66}$$

We have thus shown the following: performing the first operation in (1.62) together with a Lorentz transformation on the independent variables gives the same result as performing a Lorentz transformation on the vector field m^μ. The analogous statement holds for n^ν. Consequently the operations entering the Maxwell equations are Lorentz invariant. Therefore the Maxwell equations themselves are also Lorentz invariant provided the object given by

$$j^\mu = \begin{pmatrix} \rho \\ \mathbf{j} \end{pmatrix}$$

is considered a 4-vector field. We will later describe the much more natural description of the Lorentz invariance of electromagnetism due to Minkowski.
It is known from Einstein's writings that he was from early on fascinated by the symmetry between electric and magnetic phenomena: "boosting a purely electric field creates a magnetic field and vice versa". And indeed: this symmetry disappears in the nonrelativistic limit. There are in fact two different ways to let c go to infinity. One called "magnetic limit" consists of taking the limit directly in the above equations.

$$\nabla \mathbf{E} = 4\pi\rho \qquad -\nabla \times \mathbf{B} = 4\pi \mathbf{j} \tag{1.67}$$

and

$$\nabla \mathbf{B} = 0 \qquad -\partial_t \mathbf{B} - \nabla \times \mathbf{E} = 0 \tag{1.68}$$

This is a Galilean invariant theory with Galilei transformations acting as follows

$$\bar{\mathbf{E}} = \mathbf{E} + \mathbf{v} \times \mathbf{B} \qquad \bar{\mathbf{B}} = \mathbf{B} \tag{1.69}$$

together with

$$\bar{\rho} = \rho - \mathbf{v}\mathbf{j}, \qquad \bar{\mathbf{j}} = \mathbf{j} \tag{1.70}$$

The other choice called "electric limit" consists of assuming that $\mathbf{B}$ is smaller than $\mathbf{E}$ in the sense that $\mathbf{H}$ defined by $c^2\mathbf{H} = \mathbf{B}$ and $\mathbf{k} = c^2\mathbf{j}$ have finite limits as c goes to infinity. The result is

$$\nabla \mathbf{E} = 4\pi\rho \qquad -\partial_t \mathbf{E} + \nabla \times \mathbf{H} = 4\pi \mathbf{k} \tag{1.71}$$

and

$$\nabla \mathbf{H} = 0 \qquad \nabla \times \mathbf{E} = 0 \tag{1.72}$$

Here Galilei transformations are given by

$$\bar{\mathbf{E}} = \mathbf{E} \qquad \bar{\mathbf{H}} = \mathbf{H} - \mathbf{v} \times \mathbf{E} \tag{1.73}$$

together with

$$\bar{\rho} = \rho \qquad \bar{\mathbf{k}} = \mathbf{k} - \mathbf{v}\rho \tag{1.74}$$

The magnetic limit is exactly the theory of electromagnetism without Maxwell's "displacement current", the electric limit that without the Faraday induction term. None of those theories is able to describe radiation.

1.6 The Galilei spacetime M_G

We have by now gotten used to the picture of physical phenomena taking place in spacetime, a 4 dimensional vector space, consisting of "events". Before the advent of relativity theory people did not spend a thought on the further structures this space might have, since they believed them to be 'self-evident'. This further structure says first of all that events are 'stacked' into three dimensional hyperplanes consisting of events which are considered as 'happening simultaneously'. In addition the amount of time elapsed between different such hyperplanes is well defined. Viewed merely as a 4 dimensional affine space our spacetime would have natural coordinates (x^0, x^1, x^2, x^3) which are unique up to transformations of the form $x^\mu \mapsto \bar{x}^\mu = A^\mu{}_\nu x^\nu + d^\mu$, with $A^\mu{}_\nu$ and d^μ constants. Now, with the additional structure, there is a natural choice of x^0 coordinate, labelling simultaneous events, which we call t. It is unique up to translations. Consequently, if we stick to this choice, the freedom of performing linear transformations gets reduced to those $A^\mu{}_\nu$'s for which $A^0{}_i = 0$ (with $i, j = 1, 2, 3$) and $A^0{}_0 = 1$. In more invariant terms we can state this as follows: there is a covector field τ_μ, given by $\tau_\mu = \partial_\mu t$, and $A^\mu{}_\nu \tau_\mu = \tau_\nu$. Next we require there to be a Euclidean metric $h_{ij} = \delta_{ij}$ (with inverse $h^{ij} = \delta^{ij}$) on each 'slice' of constant t. We can then restrict the coordinate freedom to symmetries of this Euclidean metric. This leaves us with those $A^i{}_j$'s which are rotations, i.e. satisfy $A^i{}_k A^j{}_l \delta^{kl} = \delta^{ij}$. We thus end up with $10 = 3 + 3 + 4$ parameters, i.e. 3 for rotations, 3 for $A^i{}_0$ and 4 for translations in space and time. The second '3' are of course Galilean boosts.

Exercise 16: Show that the Galilei group consists of exactly those transformations of the form $x^\mu \mapsto \bar{x}^\mu = A^\mu{}_\nu x^\nu + d^\mu$, for which $A^\mu{}_\nu A^\rho{}_\sigma h^{\nu\sigma} = h^{\mu\rho}$, where

$$h^{\mu\nu} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

and $A^\mu{}_\nu \tau_\mu = \tau_\nu$, where $\tau_\mu = (1, 0, 0, 0)$. e.o.e

We next consider motions of point particle ('observers'). These are described by paths (or "worldlines") $\lambda \in \mathbb{R} \mapsto x^\mu(\lambda) \in M_G$. There are 3 classes of such paths depending on the sign of $\epsilon = \frac{dx^\mu}{d\lambda}\tau_\mu = \dot{x}^\mu \tau_\mu = v^\mu \tau_\mu$. When ϵ is zero, v^μ is tangent to a $t = $ const surface, i.e. connects points having the same t -value, i.e. this path has infinite speed. When ϵ is non-zero, the path has finite speed and is future- or past-pointing depending on the sign of ϵ. We will clearly require $\epsilon > 0$. We can then choose $\lambda' = t$ as new parameter, in terms of which $u^\mu \tau_\mu = \frac{dx^\mu}{d\lambda'}\tau_\mu = 1$. Although there is a concept of finite or infinite velocity, the value (and direction) of velocity is relative. Consider the velocity 4-vectors u^μ and u'^μ of two observers at some event, both normalized, i.e. $u^\mu \tau_\mu = 1$, $u'^\mu \tau_\mu = 1$.

The vector $w^\mu = u'^\mu - u^\mu$, which satisfies $w^\mu \tau_\mu = 0$ has components $(0, w^1, w^2, w^3)$. The spatial vector $\vec{w} = (w^1, w^2, w^3)$ with $|\vec{w}| = \sqrt{w^i w^j \delta_{ij}}$ measures the relative speed of the two observers.

Chapter 2

Minkowski spacetime: the kinematics of special relativity

Now that we have, in the first chapter, learned that the theory of electromagnetism is invariant under Poincaré (i.e. translations in space and time plus Lorentz transformations) transformations, we are ready for adopting the view that perhaps all of physics (or most: gravity is an exception) has these same symmetries. We have also seen that these symmetries are nothing but the symmetries of Minkowski spacetime M, a 4-dimensional vector space endowed with the indefinite Minkowski metric $\eta^{\mu\nu}$, resp. $\eta_{\mu\nu}$. In particular we will have to recast the laws of mechanics into "relativistic" form, i.e. a form which is invariant under Poincaré transformations. But already on a purely kinematical level relativity has many interesting and surprising consequences, which are simply geometrical properties of M. To these we now turn.

2.1 Causal and spacelike vectors

We will in this section consider vectors - not vector fields. These vectors are simply elements of Minkowski space[1], viewed as a vector space. But we will often view Minkowski space as an affine space, where a vector can sit at any point, not just the origin. We will often write $u^\mu v^\nu \eta_{\mu\nu} = (u, v)$. This scalar product is Lorentz invariant in the sense that $(Lu, Lv) = (u, v)$, provided that L is a Lorentz transformation. Because $(Lu, Lv) = L^\mu{}_\nu u^\nu L^\rho{}_\sigma v^\sigma \eta_{\mu\rho} = u^\nu v^\sigma \eta_{\nu\sigma}$ since $L^\mu{}_\nu L^\rho{}_\sigma \eta_{\mu\rho} = \eta_{\nu\sigma}$.
We will sometimes, and in particular in General Relativity, write instead of the inner product $(,)$ or the metric, $\eta_{\mu\nu}$ in the case of Special Relativity, write the

[1] Following common usage, we will often speak of "Minkowski space" rather than "Minkowski spacetime".

'line element' ds^2, i.e.

$$ds^2 = \eta_{\mu\nu} dx^\mu dx^\nu \tag{2.1}$$

(see Sect.(1.3)) A non-zero vector v^μ is called causal when $v^\mu v^\nu \eta_{\mu\nu} \leq 0$. Setting $c = 1$, these are all vectors of the form $v^\mu = (v^0, v^i)$, for which $|\vec{v}|^2 \leq (v^0)^2$ and v^0 is non-zero. Depending on the sign of v^0 they are called future- or past-pointing. When equality holds, i.e. $|\vec{v}|^2 = (v^0)^2$, the vector is called "null" or "lightlike", otherwise timelike. The latter form two cones, with the origin removed. Together they form the "light cone", consisting of future and past light cone (emanating from the origin). A straight line along the light cone has $t = |\vec{x}|$ and so corresponds to a light ray. Timelike vectors form the interior of these cones. They are clearly tangent to curves corresponding to observers moving towards the future or past at subluminal velocity $\frac{|\vec{v}|}{|v^0|}$. As two special future-pointing timelike vectors consider $v^\mu = (1, 0, 0, 0)$ and $v'^\mu = (2, 1, 1, 1)$, which both have $(v, v) = -1 = (v', v')$. It is important to understand that there is no intrinsic difference between these two vectors. We can for example find a spatial rotation, which leaves v^μ invariant and sends v'^μ to $(2, \sqrt{3}, 0, 0)$. We can then find a boost in the x^1 direction, which sends $(2, \sqrt{3}, 0, 0)$ into $(1, 0, 0, 0)$, namely the one with velocity $w = \frac{\sqrt{3}}{2}$. (Why?) This is a value of roughly 86% of the speed of light. There is thus no preferred future-pointing timelike vector on the one-sheeted hyperboloid formed by $(v, v) = -1,\ v^0 > 0$, and the same is of course true for any one-sheeted hyperboloid of timelike vectors, future- or past-pointing. In particular, not only is there no timelike vector defining an absolute state of rest, there is also no concept of a timelike vector being "closer to the light cone" than some other timelike vector. For anything moving at subluminal velocity there is a frame of reference which "comoves", as exemplified by the above calculation.

Non-zero vectors which are not causal, are called spacelike vectors. They form the region outside the light cones. For spacetime dimensions larger than 1+1=2, the set of these vectors is connected.

We now come to the concept of time in Minkowski space. Consider a straight line, i.e. a curve $\lambda \in \mathbb{R} \mapsto x^\mu(\lambda) \in M$ of the form $x^\mu(\lambda) = v^\mu \lambda + c^\mu$, where v^μ and c^μ are constant. The vector v^μ is the tangent vector of the curve in the given parametrization, i.e. $\dot{x}^\mu(\lambda) = v^\mu$. The line is called future-pointing timelike, when v^μ is future-pointing timelike. The physical meaning is clearly that this is the history of a free ("inertial") particle (or observer) in subluminal motion. We have already seen that there is in Minkowski space no absolute concept of simultaneity which allows to measure elapsed time along all observers in a universal way (i.e. independently of the observer). But what is a sensible concept of "elapsed time" along a particular observer such as the one given by $x^\mu(\lambda) = v^\mu \lambda + c^\mu$? Take two points on the trajectory, say x^μ and y^μ, so that $y^\mu - x^\mu$ is a multiple of v^μ. We define the elapsed time τ for our observer as he passes from x^μ to y^μ as the quantity

$$\tau = \sqrt{-(y-x, y-x)} \tag{2.2}$$

For example, when $x^\mu = (0,0,0,0)$ and $y^\mu = (t,0,0,0)$, we have $\tau = |t|$. The advantage of the above definition is that is manifestly Poincaré invariant: it is invariant under translations, since it only depends on the difference vector $y^\mu - x^\mu$, and it is Lorentz invariant, since it only uses the invariant scalar product define by the Minkowski metric. We emphasize that the meaning of elapsed time refers to the inertial observer travelling from the event x^μ to the event y^μ. We could imagine other observers moving along a line connecting x^μ with y^μ, which is curved: no statement is for the time being made about the time elapsed for those non-inertial observers.

2.2 Simultaneity

We can now begin to formulate a concept of events which are simultaneous to events on the above world line **for the observer moving along this line**. Suppose y^μ lies to the future of x^μ, which is the same as that $y^\mu - x^\mu$ is a positive multiple of v^μ. Consider also $\mathcal{C}_-(y)$, the past light cone with vertex y^μ and $\mathcal{C}_+(x)$, the future light cone of x^μ an, finally, the set $\mathcal{C}_-(y) \cap \mathcal{C}_+(x)$. This consists of all vectors z^μ for which both $(z-y, z-y) = 0$ and $(z-x, z-x) = 0$. We now claim that these vectors (which form a two dimensional set homeomorphic to a sphere) lie in a hyperplane which is orthogonal to the world line of our inertial observer and which intersects this world line at the event corresponding to "half-time" along the observer's journey from x^μ to y^μ. The proof of this statement is very simple: We have that $(z,z) - 2(z,y) + (y,y) = 0$ as well as $(z,z) - 2(z,x) + (x,x) = 0$. Subtracting these equations from each other, we find $2(z, x-y) + (y,y) - (x,x) = 0$. The latter equation can also be written as

$$(z - \frac{x+y}{2}, y - x) = 0\,, \tag{2.3}$$

and this says that the difference between z and the vector pointing to the mid point between x and y is orthogonal to $y-x$. Since clearly the mid point between x and y is the event with elapsed time half of that between x and y, the proof is complete.

Let conversely an event lie in one of the parallel hyperplanes orthogonal to v^μ. Then we can find points x and y on the worldline, so that $z \in \mathcal{C}_-(y) \cap \mathcal{C}_+(x)$. Let z_0 be the point at which the given hyperplane intersects the worldline and write the worldline as $x^\mu(\lambda) = v^\mu \lambda + z_0^\mu$. Then the future and past light cones emanating from z intersect the worldline at values for λ given by the equation

$(z - v\lambda - z_0, z - v\lambda - z_0) = 0$. Since $(z - z_0, v) = 0$, this can be solved by

$$\lambda_\pm = \pm \left[-\frac{(z - z_0, z - z_0)}{(v, v)} \right]^{\frac{1}{2}} \tag{2.4}$$

The above square root makes sense: (v, v) is negative and $v - v_0$ is a spacelike vector, i.e. has positive Minkowski norm, since it is orthogonal to a timelike vector, namely the vector v. To prove the last implication it is simplest to take the case $v^\mu = (v^0, 0, 0, 0)$. Then all orthogonal vectors are of the form $(0, z^1, z^2, z^3)$, which are clearly spacelike. Since $z_+ = z_0 + \lambda_+ v$ lies on $\mathcal{C}_+(z)$ and $z_- = z_0 + \lambda_- v$ lies an $\mathcal{C}_-(z)$, z lies on $\mathcal{C}_-(z_+) \cap \mathcal{C}_+(z_-)$, and we are done.
Now the union of the straight line which in the above construction connects z_- and z and that connecting z with z_+ can be interpreted as a light ray emitted by the observer at z_-, reflected at z and reabsorbed by the observer at z_+. The point z_0 corresponds to half-time between emission and absorption. It is natural to regard all such points z as events which, from the viewpoint of the given observer, are simultaneous with the event z_0. As we have seen, they can alternatively be characterized as forming the hyperplane through z_0 which is orthogonal to the worldline of the observer.

Exercise 17: Using $\eta_{\mu\nu}$ we can "lower indices" on contravariant objects. For example for any vector a^μ we can define a covector ω_μ by $\omega_\mu = \eta_{\mu\nu} a^\nu$. We can also raise indices on covariant objects using $\eta^{\mu\nu}$. This leads back to the original object. Consequently no inconsistency arises when we keep the the name of an object when (some of) its indices are raised or lowered, e.g. $a_\mu = \eta_{\mu\nu} a^\nu$. Let v be a timelike vector. Consider the mixed tensor given by

$$\Pi^\mu{}_\nu(v) = \delta^\mu{}_\nu - \frac{1}{(v, v)} v^\mu v_\nu \tag{2.5}$$

Show that $\Pi^\mu{}_\nu \Pi^\nu{}_\rho = \Pi^\mu{}_\rho$, i.e. Π is a projection operator. Next show, that Π annihilates v and projects onto the subspace of vectors orthogonal to v. Then show that Π is self-adjoint w.r. to $(.\,,.)$, i.e. $(\Pi\, a, b) = (a, \Pi\, b)$ for all vectors a and b.

2.3 The reverse triangle inequality and the twin 'paradox'

A fundamental inequality in Euclidean geometry is the Cauchy-Schwarz inequality. Remarkably, there is a similar inequality in Minkowski space, but which goes the other way, provided one of the vectors is causal. It says that

$$(v, w)^2 \geq (v, v)(w, w) \tag{2.6}$$

provided that v (or w) satisfies $(v, v) \leq 0$. The simplest proof goes as follows. Suppose first that v is timelike. Then, after a suitable Lorentz transformation, it can be written as $v^\mu = (v^0, 0, 0, 0)$. The l.h. side of (2.6) is then simply $(v^0 w^0)^2$, whereas the r.h. side is given by $-(v^0)^2[-(w^0)^2 + |\vec{w}|^2]$, which is clearly less or equal. Since null vectors can be gotten as limits of timelike vectors, the statement for causal vectors follows by continuity of $(v, w)^2 - (v, v)(w, w)$ with respect to the vector v, and the proof is complete.
Suppose conversely that v is timelike and equality holds in (2.6). Then by a similar argument it follows that w has to be poportional to v. We state without proof that, when v and w are both causal and equality holds in (2.6), then again v and w have to be proportional.
Let now a and b be causal vectors, both future-pointing. Then there sum is also future-pointing. To see this we first observe that two future-pointing causal vectors have $(a, b) \leq 0$. (Proof: when a is timelike, go to a frame where $a^\mu = (a^0, 0, 0, 0)$ and use that $a^0 > 0$ and $b^0 > 0$. For causal vector a again use continuity.) Now compute $(a + b, a + b) = (a, a) + 2(a, b) + (b, b)$, and this is non-positive. Furthermore $a^0 + b^0 > 0$. This proves our claim.
We will now show that

$$[-(a + b, a + b)]^{\frac{1}{2}} \geq [-(a, a)]^{\frac{1}{2}} + [-(b, b)]^{\frac{1}{2}} \,, \tag{2.7}$$

which is the content of the reverse triangle inequality.
Proof: Since $a+b$ is causal, the l.h. side makes sense. We take the square of both sides of (2.7). After cancelling the terms (a, a) and (b, b) on both sides we find that the inequality to be proved is the same as $|(a, b)| = -(a, b) \geq [(a, a)(b, b)]^{\frac{1}{2}}$, and this is true by virtue of the reverse CS-inequality. Again equality can only hold when a and b are proportional.
Let us see what the reverse triangle inequality says physically. Suppose we have an inertial observer, say T_1, travelling from x to y, where y lies to the future of x in the sense that the vector $y - x$ is future-pointing timelike (or y lies inside $\mathcal{C}_+(x)$). Now take another observer, say T_2 who meets T_1 in x, after some time, at the event w say, suddenly stops and travels in the reverse direction to finally meet T_1 in the event y. Now consider the future-pointing timelike vectors $a = w - x$ and $b = y - w$. From the reverse triangle inequality we infer that

$$[-(a, a)]^{\frac{1}{2}} + [-(b, b)]^{\frac{1}{2}} < [-(a + b, a + b)]^{\frac{1}{2}} = [-(y - x, y - x)]^{\frac{1}{2}} \,. \tag{2.8}$$

Now the l.h. side of (2.8) is the sum of the times it took T_2 to travel from x to w and from w to y. The r.h. side is the time it took T_1 to travel from x to y. This is the essence of the famous twin 'paradox': at the reunion in the event y the twin T_1 has aged faster than the twin T_2. We will later mention a generalization of this statement in the form of a maximum principle, according to which the straight line form x to y is, in the sense of elapsed time, the longest among all curves which correspond to arbitrarily accelerated observers moving from x to y.

Another extremum principle which runs counter to Euclidean intuition is the following. Suppose we have a timelike straight line $x(\lambda)$, say through the origin, and an event z lying outside this line. We consider the quantity given by $D^2(\lambda) := (z-x(\lambda), z-x(\lambda))$. Then there is an event x_0 on $x(\lambda)$, which maximizes D^2, and this is given by the intersection with $x(\lambda)$ of the hyperplane through z, which is orthogonal to $u(\lambda) = \dot{x}(\lambda)$. For the proof we take the derivative of $D^2(\lambda)$ and set it zero. Since $x(\lambda) = u\lambda$, there is a unique solution λ_0 given by $\lambda_0 = -\frac{(x,u)}{(u,u)}$. We can check that $z - x(\lambda_0)$ is orthogonal to u. Furthermore the second derivative of D^2 is everywhere negative. This proves our claim.
It follows that $(z - x(\lambda_0), z - x(\lambda_0))$ is positive, so λ_0 also maximizes $D(\lambda) := \sqrt{D^2(\lambda)}$. This statement can be phrased as "a rod is longest in its own rest system". We will later return to this phenomenon of 'length contraction'.

2.4 Time dilation and gamma factor

A racing cyclist, after having taken Epo, drives along a linear, 60 km long race course at 60 km/hr. How long does this take him? The obvious answer, namely $T = 60$ minutes, is only true if this refers to the reading of the official stop watch at his arrival. The reading τ of his own watch will be slightly less. To calculate it, consider three events: the event of his departure from the start, which we take to be the origin without loss, the event d corresponding to the stop watch at the finish being started, and the event y of the stop watch at the finish being stopped, which is the cyclist's arrival at the finish. We write $y = d + u$. Now assume that the stop watch is started at the time of his departure - **in the common rest system of start and finish**. (One might argue that it takes at least light travel time in order for people at the finish to learn that the race has been started, but that is not the issue here: knowing in advance by whatever means the spatial distance D between start and finish in their common rest system, one could always correct for this time delay.) This assumption means that the vector d is orthogonal to the vector u, i.e. $(d, u) = 0$. It follows that $(y, y) = (d, d) + (u, u)$, what is the same as $-\tau^2 = D^2 - T^2$ or $\tau/T = \sqrt{1 - V^2} \sim 1 - \frac{V^2}{2}$ with V being the cyclist's speed in units of the speed of light. Thus $\frac{T-\tau}{T}$ is of the order of 10^{-15} in the present case. If the cyclist would turn around at the finish and cycle back, and the stop watch had been placed at the start, the above effect is just the twin paradox described in the previous section. The effect of time dilation is for example seen in cosmic ray showers, where, for the given speed, certain particles would have otherwise decayed by the time they hit the earth surface.
The ratio T/τ is of course the gamma factor encountered previously. We can use the above discussion to give a beautiful intrinsic definition of the gamma factor (equivalently: relative speed), as follows. Consider two arbitrary timelike vectors u and v. These describe the motion of two inertial observers. We claim that the

absolute value of their relative speed $V = |V|$ can be written as

$$V^2 = 1 - \frac{(u,u)(v,v)}{(u,v)^2} \tag{2.9}$$

By the reverse CS inequality we have $0 \leq V < 1$ with $V = 0$ exactly when u and v are proportional. To prove (2.9) we merely take for u and v the y and u in the above discussion, where we had $V^2 = -\frac{(d,d)}{(u,u)}$. Using that $d = y - u$ and $(u,u) = (u,y)$, this can be written in the form (2.9). The point is that the form (2.9) is invariant under independent rescalings of both u and v.
It now follows that, when u and v are future-pointing, the gamma factor can be written as

$$-\gamma(V) = \frac{(u,v)}{\sqrt{(u,u)(v,v)}}\,. \tag{2.10}$$

It is thus an analogue of the concept of (cosine of) angle in Euclidean geometry.

Exercise 18: Observer u' moves at speed V in the x^1- direction relative to observer u, and u'' moves at the same speed relative to u' and in the same direction. Show that the speed W of u'' relative to observer u is given by $W = \frac{2V}{1+V^2}$. What is the general formula?

2.5 Proper time

We now consider worldlines $z^\mu(\lambda)$ which not necessarily straight lines. Let $u^\mu(\lambda) = \dot{z}^\mu(\lambda)$ be the tangent vector. This should be pictured, for each value of λ, as a vector sitting at the point $z^\mu(\lambda)$ of Minkowski space. The worldline is called future-pointing timelike, if the vector u^μ is future-pointing timelike. The geometrical picture should be that there is a future-light cone sitting at every point of Minkowski space, and the curve is such that its tangent at each point should point inside this light cone. The physical meaning of such worldlines is of course that they correspond to particles in arbitrary - but subluminal - motion. One can, for such worldlines define a notion of elapsed time, called "proper time", which infinitesimally coincides with the previous one for straight lines in (2.2). Symbolically we write $ds = \sqrt{-\eta_{\mu\nu}dx^\mu dx^\nu}$. More precisely: the proper time s, along the given curve, between the events $x^\mu = z^\mu(\lambda_1)$ and $y^\mu = z^\mu(\lambda_2)$ is defined by

$$s = \int_{\lambda_1}^{\lambda_2} \sqrt{-(\dot{z}(\lambda), \dot{z}(\lambda))}\; d\lambda\,. \tag{2.11}$$

This definition, which clearly depends on the events x and y and the curve $z(\lambda)$ connecting them, does however not depend on the choice of parametrization of that curve. After setting $\lambda = F(\bar{\lambda})$ (with $F' > 0$ to keep the future-pointing nature of the curve), we see that F' drops out. Proper time is the Lorentzian

analogue of arc length in Euclidean geometry. It is an assumption in special relativity that "good" clocks record proper time.
We can use

$$s(\lambda) = \int_{\lambda_1}^{\lambda} \sqrt{-(\dot{z}(\lambda'), \dot{z}(\lambda'))}\; d\lambda' \,. \tag{2.12}$$

for defining, instead of λ, a new parameter which is unique up to an additive constant. In this new parametrization we have that

$$\frac{dz^\mu}{ds} = \frac{1}{\sqrt{-(\frac{dz}{d\lambda}, \frac{dz}{d\lambda})}} \frac{dz^\mu}{d\lambda} \tag{2.13}$$

Thus the tangent vector in the proper time parametrization has Minkowski norm equal to -1. It is easily seen that the converse is also true. The proper time normalization $(\dot{z}, \dot{z}) = -1$ is the relativistic analogue of the normalization $u^\mu \tau_\mu = 1$ in Galilean physics.
Another - non-invariant - parametrization of a timelike curve is that by inertial time relative to some Minkowskian coordinate system, which is nothing but proper time for an inertial observer associated with that system. When $z^\mu(\lambda)$ be the timelike curve one chooses $t(\lambda) = z^0(\lambda)$ as new parameter. This is possible, since $\frac{dz^0}{d\lambda}$ is positive by the future-pointing timelike nature of the curve. With this choice the tangent vector v has the form $v^\mu = (1, \vec{v})$. The spatial vector $\vec{v}$ is the spatial velocity of the curve seen by the observer at rest in the given Minkowskian frame (who, in proper time normalization, has the velocity $(1, 0, 0, 0)$). To relate v^μ to u^μ, the four-velocity w.r. to proper time corresponding to the given curve, we use (2.12) with $\lambda = t$. The result is

$$s(t) = \int_{t_0}^{t} \sqrt{1 - \vec{v}^2(t')}\; dt' \tag{2.14}$$

Thus

$$u^\mu = \begin{pmatrix} \frac{1}{\sqrt{1-\vec{v}^2}} \\ \frac{\vec{v}}{\sqrt{1-\vec{v}^2}} \end{pmatrix}$$

where the t - argument in $\vec{v}$ has to be eliminated in terms of s using the inverse of the transformation in (2.14). Of course $(u, u) = -1$, as it has to be.
En passant Eq.(2.14) shows the important fact that, for traveling between two timelike-related events, the inertial motion maximizes proper time.
Exercise 19: Make sure that the curve $(z^0(t), z^i(t)) = (t, R\cos\Omega t, R\sin\Omega t, 0)$ is timelike and calculate the proper time per period. e.o.e.

2.6 Length contraction

As in Galilean physics 'length' means 'distance between simultaneous events'. But contrary to Galilean physics, simultaneity is relative, and so therefore is

length.
Consider a rod in uniform motion. The ends of this rod are thus described by two parallel timelike lines, say $z_1^\mu(\sigma) = u^\mu\sigma$ and $z_2^\mu(\tau) = u^\mu\tau + d^\mu$. We have $(u,u) = -1$, so σ and τ are proper times. Possibly after adding a constant to τ, we can arrange for $(d,u) = 0$. The length D of the rod in its own rest system is thus given by $D^2 = (d,d)$. Now consider another observer with four velocity u'^μ which meets the z_1-end of the rod in the event corresponding to the origin. The length of the rod D' according to this observer is given by the norm of the spacelike vector d' pointing from the origin to z_2 with the property that $(d',u') = 0$. After a trivial calculation we see [2] that $d' = d - \frac{(d,u')}{(u,u')}\,u$, implying

$$D'^2 = D^2 - \frac{(d,u')^2}{(u,u')^2} \tag{2.15}$$

We consider two extreme cases:
(a) $(d,u') = 0$. This means that the primed observer, from the standpoint of the unprimed observer, moves in a spatial direction orthogonal to that of the rod. In this case (2.15) shows 'there is no effect', i.e. the primed observer measures the same length as the unprimed one.
(b) The relative motion of the observers is along the direction of the rod: geometrically this means that u, u', d are linearly dependent. Assume without loss that the motion is along the x^1-direction. In the rest system of the unprimed observer we have $u = (1,0,0,0)$, $u' = \gamma(V)(1,V,0,0)$ and $d = (0,D,0,0)$, so that $(u',d)^2 = \frac{V^2D^2}{1-V^2}$. (Invariantly this says that $(d,u')^2 = (d,d)[(u,u')^2 - 1]$.) Inserting this into (2.15) there results

$$D' = \frac{D}{\gamma} < D \tag{2.16}$$

Exercise 20: A car of 4m length is driven at speed $\gamma = 2$ into a 2m long garage. When the front end of the car hits the back wall of the garage, the rear end of the car, due to relativistic length contraction, is just inside the garage, and the garage keeper can close the door. True or false?

2.7 *Light: wave fronts and rays

Geometrical optics is an approximation to the theory of electromagnetism, or in fact any theory decribed by wavelike equations, which holds in the limit, where the "phase varies much more rapidly than the amplitude" and where, in addition, polarization effect are ignored. To get an idea how this limit is arrived at, it suffices to consider the wave equation

$$\eta^{\mu\nu}\partial_\mu\partial_\nu\Phi = 0 \tag{2.17}$$

[2] This is of course nothing but $d'^\mu = \Pi^\mu{}_\nu(u)d^\nu$ with $\Pi^\mu{}_\nu$ defined in (2.5).

We consider solutions of (2.17) with a small parameter ϵ of the form

$$\Phi(\epsilon; x) = A(\epsilon; x)\, e^{\frac{i}{\epsilon}S(x)} \tag{2.18}$$

where $A(\epsilon; x)$ is assumed to have a smooth non-zero limit as ϵ goes to zero. Inserting (2.18) into (2.17), multiplying by ϵ and letting $\epsilon \to 0$, we obtain the nonlinear equation

$$\eta^{\mu\nu}(\partial_\mu S)(\partial_\nu S) = 0 \tag{2.19}$$

Level sets of a function $S(x)$ (i.e. the hypersurfaces $S(x) = \text{const}$) satisfying (2.19) are called wave fronts. Here are two examples:
(a) $S(x) = k_\mu x^\mu$, where k is a constant covector with $\eta^{\mu\nu}k_\mu k_\nu = 0$. Associated wave fronts are hyperplanes. These describe plane waves $\Phi \sim e^{\omega t + \vec{k}\vec{x}}$ with $|\omega| = |\vec{k}|$.
(b) $S(x) = \eta_{\mu\nu}x^\mu x^\nu$. This solves (2.19), but only at points x for which $S(x)$ vanishes. The associated wave front $\{x \in M | S(x) = 0\}$ is the light cone through the origin. The functions $\Phi(x) \sim e^{i(x,x)}$, which are singular at the origin, describe a spherical wave contracting towards the origin and re-expanding.
Define $k_\mu(x) = \partial_\mu S(x)$, where S solves (2.19). Generally the gradient of a function S is called a co-normal of the hypersurface given $S(x) = \text{const}$. Its meaning is that vectors v at points on the hypersurface defined by $v^\mu \partial_\mu S = 0$ are exactly the tangents to curves which lie on the hypersurface. If, as we assume, $\partial_\mu S$ is not the zero-vector, this is a 3-dimensional set of vectors at each point of the hypersurface. Next consider the 'normal' vector k^μ given by $k^\mu = \eta^{\mu\nu}\partial_\nu S$. By (2.19), this is a vector tangent to the wave front. So the normal of a wave front is tangent to the wave front, rather than pointing out of it - an impossible situation in Euclidean geometry[3]! And thus wave fronts, or null hypersurfaces, are 3 dimensional surfaces in spacetime for which three tangent vectors k, a, b can be chosen so that $(k, k) = 0 = (k, a) = (k, b)$. We mention without proof that the opposite is also true: for every null surface there is a function $S(x)$ satisfying (2.19), so that this null surface is given by $S(x) = \text{const}$.

Exercise 21: Show that a vector which is orthogonal to a null vector is either proportional to this null vector or spacelike. Hint: Show first that by a suitable Lorentz transformation every null vector can be transformed to the form $k^\mu = (k^0, k^1, 0, 0)$. e.o.e.

A fundamental property of wave fronts is that the normal vector field k^μ is tangent to straight lines. To show this consider the vector field $k^\nu \partial_\nu k^\mu = \eta^{\mu\rho}k^\nu \partial_\nu k_\rho$. We have

$$\eta^{\mu\rho}\eta^{\nu\sigma}\partial_\sigma S(\partial_\nu\partial_\rho S) = \eta^{\mu\rho}\eta^{\nu\sigma}\partial_\sigma S(\partial_\rho\partial_\nu S) = \frac{1}{2}\eta^{\mu\rho}\partial_\rho[\eta^{\nu\sigma}(\partial_\nu S)(\partial_\sigma S)] = 0 \tag{2.20}$$

[3]and one of many reasons for carefully distinguishing between vectors and covectors.

Thus $k^\nu \partial_\nu k^\mu$ is zero. But this says that k^μ, viewed as the tangent vector to the the curve $x^\mu(\lambda)$, satisfies $\dot{x}^\nu \partial_\nu k^\mu = \frac{d}{d\lambda} k^\mu = \ddot{x}^\mu = 0$, i.e. the curve is a straight line. Consequently wave fronts are hypersurfaces, which are 'ruled by' null rays. We finally outline how wave fronts can be constructed. One starts with an 'initial wave front': this is a 2-surface $x^\mu = x^\mu(y^A)$, $A = 1, 2$. We seek a wave front going through that surface, representing the time evolution of that initial surface. Take for example the 2-surface given by $x^\mu = (0, x^i(y^A))$. At each point of this surface we seek vectors $k^\mu(y^A)$ which are future-pointing null and normal to the initial surface, and which serve as initial data for $k^\nu \partial_\nu k^\mu = 0$. These can, up to a factor, be written as $k^\mu = (1, \vec{n})$, where $\vec{n}^2 = 1$, where $\vec{x}_{,1}\vec{n}$ and $\vec{x}_{,2}\vec{n}$ should both be zero. These equations determine $\vec{n}(y^A)$ up to sign. The two null vector fields $k^\mu_\pm(y) = (1, \pm\vec{n}(y))$ are the sought for initial data, they satisfy $x^\mu_{,A} k^\nu \eta_{\mu\nu} = 0$. Thus we obtain two wave fronts emanating from the given initial wave front, one 'outgoing', one 'ingoing'. Take for example the initial wave front given by a sphere of radius R in the $t = 0$ surface. i.e. $x^\mu(\theta, \phi) = (0, R\sin\theta\cos\phi, R\sin\theta\sin\phi, R\cos\theta)$. The two null vector fields $k^\mu_\pm(y)$ can then be written as $k^\mu_\pm(\theta, \phi) = (1, \pm\sin\theta\cos\phi, \pm\sin\theta\cos\phi, \pm\cos\theta)$. The associated null surfaces are simply the null cones emanating from the points $z^\mu_\pm = (\mp R, 0, 0, 0)$.

2.8 *Phase and frequency

We now explain the subtle concept of phase for bundles of light rays. Suppose we have a 2 parameter set of events $x^\mu = x^\mu(\sigma, \tau)$ which, for fixed τ, describes light rays and, for fixed σ, timelike curves. Consider two such curves, say γ_e and γ_a, intersecting the null rays transversally and with the same orientation, perhaps but not necessarily given by $\sigma = \sigma_1$ and $\sigma = \sigma_2$. Let $k_\mu = \eta_{\mu\nu} \frac{\partial x^\mu}{\partial \sigma}$. We claim that

$$\int_{\gamma_e} k_\mu dx^\mu = \int_{\gamma_a} k_\mu dx^\mu \,, \tag{2.21}$$

where the integral on both sides is taking between τ_1 and τ_2, that is to say between the intersection of the two transversal curves with two fixed null rays.
For the proof, we take the difference between the two sides in (2.21) and complete it to closed line integral of $\int k_\mu dx^\mu$ by adding the appropriate segments along null rays. These however are clearly zero since there $dx^\mu = k^\mu d\sigma$ and $k_\mu k^\mu$ vanishes. We are thus left with showing that

$$\oint k_\mu dx^\mu = 0 \tag{2.22}$$

for all closed curves in the space spanned by σ and τ. Now recall the Stokes theorem in the plane which can be written as

$$\oint_\gamma (Pdx + Qdy) = \iint_\Omega \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dxdy \,, \tag{2.23}$$

where Ω is a bounded open domain bounded by the closed curve γ. There is an analogous formula holding for arbitrary (not necessarily planar) 2 dimensional surfaces in $\mathbb{R}^n$ which says

$$\oint k_\mu dx^\mu = \iint (\partial_\mu k_\nu - \partial_\nu k_\mu) \frac{\partial x^\mu}{\partial \sigma} \frac{\partial x^\nu}{\partial \tau} d\sigma d\tau \tag{2.24}$$

Note that (2.24) reduces to (2.23) if we put $P(\sigma,\tau) = k_\mu(\sigma,\tau)\frac{\partial x^\mu}{\partial \sigma}$ and $Q(\sigma,\tau) = k_\mu(\sigma,\tau)\frac{\partial x^\mu}{\partial \tau}$, as a little calculation shows[4].
To prove (2.22) consider $(\partial_\mu k_\nu - \partial_\nu k_\mu)k^\mu$. Here the first term is zero, since $\frac{\partial^2 x^\nu}{\partial \sigma^2} = 0$. And the second term is zero since it is equal to $-\frac{1}{2}\,\partial_\nu(k_\mu k^\mu)$. Combining this with (2.24) ends the proof of (2.22).
Now to the concept of frequency. This is defined as 'change of phase w.r. to proper time' of an observer intercepting the given bundle of rays, i.e.

$$\lim \frac{\Delta S}{\Delta s} = \lim \frac{1}{\Delta s} \int k_\mu dx^\mu = \lim \frac{1}{\Delta s} \int k_\mu u^\mu ds = k_\mu u^\mu \tag{2.25}$$

where $(u,u) = -1$. Since k_μ is parallelly transported along the ray, we can compute the frequency along some light ray for two observers with unit tangents u_1 and u_2 at any single point on the ray. Strictly speaking only relative frequencies make sense, i.e.

$$\frac{\omega_1}{\omega_2} = \frac{(k,u_1)}{(k,u_2)} \tag{2.26}$$

This has the following reason: A light ray is governed by the equation $k^\nu \partial_\nu k^\mu = 0$ together with $(k,k) = 0$. If we replace k by any constant multiple, it will still be tangent to the same ray and satisfy $k^\nu \partial_\nu k^\mu = 0$, Thus the concept of phase as described has an inherent dependence on a choice of scale for k which drops out, when relative frequencies are considered.

2.9 Doppler effect

Take first the case where u_1, u_2 and k are linearly dependent, i.e. everything happens in one spatial direction, say in the 2-dimensional Minkowski space spanned by the coordinates t and x^1. Let for example the light ray move to the right and our 3 vectors be oriented, from left to right, as u_1, u_2 and k. Thus, from the viewpoint of u_1, u_2 moves after the light ray. And, from the viewpoint of u_2, u_1 moves against the light ray. We thus expect $\omega_1 > \omega_2$. We can set $u_1 = (1,0)$ and $u_2 = \gamma(V)(1,V)$, where $V > 0$. Finally we can set $k = (1,1)$. In more invariant terms, and for later use, these relations can be written as

$$u_2 = \sqrt{\frac{1-V}{1+V}}\, u_1 - \frac{V}{\sqrt{1-V^2}} \frac{1}{(k,u_1)}\, k \tag{2.27}$$

[4]Hint: 2nd derivatives of $x^\mu(\sigma,\tau)$ cancel.

Using this, or from the explicit choice of our three vectors we find

$$\frac{(k,u_1)}{(k,u_2)} = \frac{\omega_1}{\omega_2} = \frac{-1}{\gamma(V)(-1+V)} = \sqrt{\frac{1+V}{1-V}} \tag{2.28}$$

At the other extreme consider the situation where u_2, from the viewpoint of u_1 moves orthogonally to the light ray, in other words we assume that $\Pi^{\mu}{}_{\nu}(u_1)u_2^{\nu}$ is orthogonal to k. Then

$$\frac{(k,u_2)}{(k,u_1)} = -(u_1,u_2) = \frac{1}{\sqrt{1-V^2}} \tag{2.29}$$

Exercise 22: Relative to some inertial frame a source of proper (i.e. with respect to its own rest system) frequency ω_1 and an observer move along coplanar straight lines and with the same speed V. A light ray travels from source to observer along a (spatial) line l. At emission the source crosses l at 30° 'to the right' while at reception the observer crosses l at 30° 'to the left' (for someone looking along l). Prove the observed frequency ω_2 is given by $\omega_2 = \frac{2-\sqrt{3}V}{2+\sqrt{3}V}\,\omega_1$. e.o.e.
Exercise 23: Let, in the reference frame of u_1, the light ray and u_2 enclose the angle α. Then show that

$$\frac{\omega_2}{\omega_1} = \frac{1-|V|\cos\alpha}{\sqrt{1-V^2}} \tag{2.30}$$

Exercise 24: The 'drag effect': Light propagation in a moving medium (in $1+1$ dimensions) is governed by the contravariant metric

$$g^{\mu\nu} = \eta^{\mu\nu} + u^{\mu}u^{\nu}(1-\frac{1}{n^2})\,, \tag{2.31}$$

where $n<1$ is the index of refraction and u^{μ} is the four (in our case: 'two'-) velocity vector of the medium. The phase velocity $\bar{c}$ of light - propagating to the right,say - is then given by $g^{\mu\nu}k_{\mu}k_{\nu}=0$, where $k_{\mu}=(1,\frac{1}{\bar{c}})$. Let the medium move in some inertial frame at velocity $\pm V$ with $0<V<\frac{1}{n}$. Show that

$$\bar{c} = \frac{\frac{1}{n}\pm V}{1\pm V\frac{1}{n}} \tag{2.32}$$

and interpret the result. e.o.e.

2.10 Aberration of light

Consider two future-pointing null rays k and l as will as two observers u and u', i.e. $(u,u)=(u',u')=-1$, both future-pointing. The spatial angle between l and k depends on the observer. For u it is given by

$$\cos\Theta = \frac{(\Pi k,\Pi l)}{\sqrt{(\Pi k,\Pi k)(\Pi l,\Pi l)}} \tag{2.33}$$

where the projection Π is along the vector u. After a little calculation this leads to

$$\cos\Theta - 1 = \frac{(k,l)}{(k,u)(l,u)} \tag{2.34}$$

and of course

$$\cos\Theta' - 1 = \frac{(k,l)}{(k,u')(l,u')} \tag{2.35}$$

We now assume that the vectors k, u, u' are coplanar in the same configuration as k, u_1 and u_2 in the previous section. It then follows from taking the scalar product of (2.27) with l that

$$\begin{aligned}(l,u') &= \sqrt{\frac{1-V}{1+V}}(l,u) - \frac{V}{\sqrt{1-V^2}}\frac{(l,k)}{(k,u)} \\ &= \sqrt{\frac{1-V}{1+V}}(l,u) - \frac{V}{\sqrt{1-V^2}}(\cos\Theta - 1)(l,u) \ ,\end{aligned} \tag{2.36}$$

where we have used (2.34) in the second line. Taking the ratio of (2.34) and (2.35) and using (2.36, 2.28), a short calculation shows

$$\frac{\cos\Theta' - 1}{\cos\Theta - 1} = \frac{1+V}{1-V\cos\Theta} \tag{2.37}$$

or equivalently

$$\cos\Theta' = \frac{\cos\Theta - V}{1 - V\cos\Theta} \tag{2.38}$$

Note that, since $V > 0$ we have for $0 < \Theta < \pi$ that $\cos\Theta > \cos\Theta'$. What does this mean? Suppose we are in the rest system of u. Then u' moves 'after' the light ray given by k, so that, for u, the angle between k and l is the same as that between l and u'. Furthermore, from above, there holds $\Theta' > \Theta$. If light rays are replaced by subluminal particles, this phenomenon is known by cyclists: if there is rain falling vertically ($\Theta = \frac{\pi}{2}$) in some rest system, it appears to come from the forward direction in the moving system (i.e. $\Theta' > \frac{\pi}{2}$).

Exercise 25: An astronomer observes the angles $\Theta_{12}, \Theta_{13}, \Theta_{34}, \Theta_{24}$ between four astronomic objects. Show that the quantity

$$\frac{(1-\cos\Theta_{12})(1-\cos\Theta_{34})}{(1-\cos\Theta_{13})(1-\cos\Theta_{24})} \tag{2.39}$$

is independent of the state of motion of the observer. e.o.e.

Chapter 3

Relativistic mechanics

3.1 Four momentum

It is a fundamental principle of mechanics that a system of (point-)particles interacting through conservative forces which only depend on relative distances has 10 conserved quantities, namely total energy, momentum, angular momentum and centre of mass. When the particles do not interact, then these quantities are conserved separately for each particle. One can now consider collisions: these are solutions to the dynamical equations wbich consist of a transition between two states given by non-interacting particles. [1] Before and after the collision the conserved quantities are simply the sum of these quantities for each individual particles. This already gives a number of constraints of possible end states given an initial state independently of the details of the interaction. We will in the following mainly consider energy and momentum conservation.
Consider an elastic collision in Newtonian mechanics, with $1 \leq i \leq k$ initial particles and $1 \leq f \leq l$ final particles. Energy conservation says that

$$\sum_i \frac{m_i}{2} \mathbf{v}_i^2 = \sum_f \frac{m_f}{2} \mathbf{v}_f^2 \tag{3.1}$$

and momentum conservation that

$$\sum_i m_i \mathbf{v}_i = \sum_f m_f \mathbf{v}_f \tag{3.2}$$

These relations have to hold in all inertial frames. Thus they should remain true after the replacement $\mathbf{v}_i \mapsto \mathbf{v}_i - \mathbf{V}$ together with $\mathbf{v}_f \mapsto \mathbf{v}_f - \mathbf{V}$ has been made. We easily see that this requires

$$\sum_i m_i = \sum_f m_f \ , \tag{3.3}$$

[1] An example is elastic scattering of two particles interacting via a potential which suitably goes to 0 at infinite distance between the particles.

i.e. mass has to be conserved. We now have to find the correct relativistic form of energy and momentum. The basic requirements are Lorentz invariance and agreement with the nonrelativistic laws for small velocities. It is suggestive to use for energy and momentum the four vector given by

$$p^\mu = mu^\mu \ , \tag{3.4}$$

where u is normalized, i.e. $(u, u) = -1$. This vector is called four momentum or energy-momentum. In some standard Minkowskian coordinate system the vector p^μ takes the form

$$p^\mu = m \begin{pmatrix} \frac{1}{\sqrt{1-\vec{V}^2}} \\ \frac{\vec{V}}{\sqrt{1-\vec{V}^2}} \end{pmatrix}$$

where $\mathbf{V}$ is the three-velocity associated with u^μ in the given system. The interpretation will be that the zero-component is the energy, the spatial components the momentum of the particle in the given coordinate system. What does this mean in more invariant terms? Recall from linear algebra that a Minkowskian coordinate system is the same as a choice of 'standard tetrad', i.e. a quadruple of vectors (v, e_1, e_2, e_3) with $(v, v) = -1$, $(v, e_i) = 0$, $(e_i, e_j) = \delta_{ij}$. If only a choice of v is made, this already defines a unique inertial system. (Then the choice of e_i is unique up to a spatial rotation w.r. to v.) Energy is now defined by

$$E = -(v, p) \tag{3.5}$$

and the components of momentum are given by

$$p_i = p^i = (p, e_i) \tag{3.6}$$

Whereas momentum clearly approximates the non-relativistic expression, this is not true for energy. Namely writing for E the expression

$$E = mc^2 \frac{1}{\sqrt{1 - \frac{\vec{V}^2}{c^2}}} \ , \tag{3.7}$$

where the insertion of the factor of c^2 gives E the dimension of energy, we find that

$$E = mc^2 \left[1 + \frac{\vec{V}^2}{2c^2} + O\left(\left(\frac{\vec{V}^2}{c^2} \right)^2 \right) \right] \sim mc^2 + \frac{m\vec{V}^2}{2} + ... \tag{3.8}$$

So the energy of a free particle is to leading order given by mc^2 plus the non-relativistic kinetic energy. The first term is for obvious reasons called the rest-energy.

It is now clear what form conservation of energy and momentum takes for a relativistic system of point particles, namely

$$\sum_i p_i{}^\mu = \sum_f p_f{}^\mu \,, \tag{3.9}$$

where $p_i{}^\mu = m_i u_i{}^\mu$ and analogously for the outgoing particles. Take for example the process where a single particle of mass M decays into several particles with masses m_f. This is impossible in a Galilean world. In the rest system of the M-particle, we have that the sum of the kinetic energies of the outgoing particles is zero. By positive definiteness of the kinetic energies, each outgoing particle has to have zero momentum. Thus the M-particle does not decay, due to to lack of available kinetic energy. Relativistically the situation changes. Suppose for simplicity that the M-particle wants to decay into two particles of equal mass m. In the rest system of the M particle we have that $P^\mu = M(1, \vec{0})$. Thus

$$\frac{\vec{V}_1}{\sqrt{1-\vec{V}_1^2}} + \frac{\vec{V}_2}{\sqrt{1-\vec{V}_2^2}} = 0 \,, \tag{3.10}$$

which is easily seen to imply that $\vec{V}_1 = -\vec{V}_2 = \vec{V}$. Conservation of energy now gives that

$$M = \frac{2\,m}{\sqrt{1-\vec{V}^2}} \tag{3.11}$$

Thus the necessary kinetic energy for the decay products is supplied by the difference between M and $2m$. In fact, from the reverse triangle inequality we know that $M > m_1 + m_2$ except when all three vectors are proportional, which would mean the process does not take place.
If, conversely, we want to form a heavier particle of mass M by shooting particles of mass m at each other, the light particles have to have velocities $\frac{\vec{V}^2}{c^2} \geq 1-4(\frac{m}{M})^2$. It turns out that this 'energy-mass-equivalence' applies to all forms of energy, e.g. binding energy U_{int}. This for example explains the fact that the mass of a stable nucleus exceeds the sum of masses of its constituents: the difference is the mass equivalent of the binding energy.

Exercise 26: Show that the the energy $E(v)$ of a massive particle with 4-momentum p relative to the observer with 4-velocity v (with $(v, v) = -1$) has the minimum value m at $v = \frac{p}{m}$. e.o.e.
Exercise 27: The center of momentum frame for a system of particles with momenta p_1, p_2, etc. is defined by a four vector u with $(u, u) = -1$ subject to the equation $mu = \sum_i p_i$. Express the total mass m in terms of the relative velocities of the particles. e.o.e.

3.2 Relativistic billiards

We are here considering the elastic scattering of two particles of equal mass. Calling initial (resp. final) four momenta p and q (resp. p' and q') our equations are

$$p + q = p' + q' \, , \tag{3.12}$$

where $(p,p) = (q,q) = (p',p') = (q',q') = -1$. We will work in the rest frame of q. Namely we will be interested in the angles $\Theta_1 = \angle(\Pi p', \Pi p)$ and $\Theta_2 = \angle(\Pi q', \Pi p)$, where $\Pi^\mu{}_\nu = \delta^\mu{}_\nu + q^\mu q_\nu$. We are here assuming that neither $\Pi p'$ nor $\Pi q'$ are zero, which would correspond to the situation where one particle, typically p' is at rest after the collision and $q' = p$. We claim that $0 < \Theta_{1,2} < \frac{\pi}{2}$ and $\tan\Theta_1 \tan\Theta_2 = \frac{2}{-(p,q)+1}$. To understand the meaning of this relation, recall the $-(p,q)$ is the gamma factor of the p particle in the rest frame of q. In the non-relativistic limit $\frac{\vec{V}^2}{c^2} \sim 0$, $\frac{2}{-(p,q)+1}$ goes to 1, and this means that $\Theta_1 + \Theta_2 = \frac{\pi}{2}$, a fact well known to nonrelativistic billiard players. In the extreme relativistic limit the gamma factor goes to infinity, so that at least one of $\Theta_{1,2}$ goes to zero. We now turn to the proof of the above statement. By straightforward computation using the definition of Π, we find

$$\Pi p = p + (p,q)q \qquad \Pi p' = p' + (p',q)q \tag{3.13}$$

$$\Pi q = 0 \qquad \Pi q' = q' + (q',q)q \tag{3.14}$$

as well as

$$(\Pi p, \Pi p) = -1 + (p,q)^2 \qquad (\Pi p', \Pi p') = -1 + (p',q)^2 \tag{3.15}$$

$$(\Pi q, \Pi q) = 0 \qquad (\Pi q', \Pi q') = -1 + (q',q)^2 \tag{3.16}$$

and

$$(\Pi p, \Pi p') = (p,p') + (p,q)(q,p') \tag{3.17}$$

Using Eq.(3.12), the relation

$$(p,q) = (p',q') \tag{3.18}$$

which follows from (3.12) and

$$(p',p) = (p', p+q-q) = (p', p'+q'-q, q) = -1 + (p,q) - (p',q) \tag{3.19}$$

we infer that

$$(\Pi p', \Pi p) = [-(p,q)+1][-(p',q)-1] \, . \tag{3.20}$$

Similarly

$$(\Pi q', \Pi p) = [-(p,q)+1][-(q',q)-1] \tag{3.21}$$

Putting all of this together, there results

$$\cos\Theta_1 = \sqrt{\frac{-(p,q)+1}{-(p,q)-1}}\sqrt{\frac{-(p',q)-1}{-(p',q)+1}} \tag{3.22}$$

$$\cos\Theta_2 = \sqrt{\frac{-(p,q)+1}{-(p,q)-1}}\sqrt{\frac{-(q',q)-1}{-(q',q)+1}} \tag{3.23}$$

From $\sin^2\Theta = 1-\cos^2\Theta$ it follows that

$$\sin^2\Theta_1 = \frac{2(p'-p,q)}{[-(p,q)-1][-(p',q)+1]} \tag{3.24}$$

Now $\cos\Theta_i > 0$. whence $0 < \Theta_i < \frac{\pi}{2}$. And since $(p'-p,q) = -1-(q',q)$, we have that $(p'-p,q) > 0$ and, similarly $(q'-p,q) > 0$. Putting all of this together it follows that

$$\sin\Theta_1 = \frac{\sqrt{2(p'-p,q)}}{\sqrt{-(p,q)-1}\sqrt{-(p',q)+1}} \tag{3.25}$$

$$\tan\Theta_1 = \frac{\sqrt{2(p'-p,q)}}{\sqrt{-(p,q)+1}\sqrt{-(p',q)-1}} \tag{3.26}$$

$$\tan\Theta_2 = \frac{\sqrt{2(q'-p,q)}}{\sqrt{-(p,q)+1}\sqrt{-(q',q)-1}} \tag{3.27}$$

Consequently

$$\tan\Theta_1\tan\Theta_2 = \frac{2}{-(p,q)+1} \tag{3.28}$$

as claimed. In the special case where the outgoing state is, in the lab frame, symmetric w.r.t. to interchange of the two particles, we have $\Theta_1 = \Theta_2 = \Theta$, so that $\tan\Theta = \sqrt{\frac{2}{-(p,q)+1}}$.

3.3 Compton effect

We are here looking at the elastic scattering of a photon at an electron. According to the quantum theory of light, a photon carries a four-momentum p^μ equal to $p^\mu = \hbar k^\mu$, where k^μ is the frequency four-vector introduced previously. Thus $p,p) = 0$. The elctron has four-momentum q^μ with $(q,q) = -m^2$, where m is the electron mass. Thus energy momentum conservation reads

$$p+q = p'+q', \tag{3.29}$$

where $(p',p') = 0$ and $(q',q') = -m^2$. Thus

$$-m^2 = (p+q-p', p+q-p') = -m^2 + 2(q, p-p') - 2(p,p') \tag{3.30}$$

or

$$(q, p - p') = (p, p')\,, \tag{3.31}$$

what is the same as

$$\hbar\, m(\omega - \omega') = \hbar^2 \omega\omega'(1 - \vec{n}\vec{n}') \tag{3.32}$$

or

$$\frac{1}{\omega'} - \frac{1}{\omega} = \frac{\hbar}{mc}\,(1 - \cos\Theta) \tag{3.33}$$

Thus the incoming electron suffers an increase of the wavelength depending on the scattering angle. The factor $\frac{\hbar}{mc}$ is the Compton wavelength of the electron.

Exercise 28 "Bremsstrahlung": A fast electron (mass m) decelerates by colliding with a stationary heavy nucleus of mass M. In the process a photon of frequency ω is emitted. Show that the maximum energy of the photon is given by

$$\hbar\,\omega_{\max} = m\,\frac{\gamma - 1}{1 + \frac{m}{M}\sqrt{\frac{1-V}{1+V}}} \tag{3.34}$$

and is attained when electron and nucleus move 'as a lump' after the collision and the photon is emitted in the forward direction. Hint: Square the equation $p + q - \hbar k = p' + q'$. e.o.e.

3.4 Accelerated motion

Let $x = z(s)$ be a timelike future-pointing curve parametrized by proper time. Then the derivative $u = \dot{z}$ satisfies $(\dot{z}, \dot{z}) = -1$. Furthermore $\dot{z}$ is related to the coordinate 3-velocity $\vec{v}$ corresponding to $z(s)$ in some inertial frame by

$$u^\mu = \gamma \begin{pmatrix} 1 \\ \vec{v} \end{pmatrix}$$

In particular

$$\dot{t} = \gamma \tag{3.35}$$

Exercise 29: A particle moves along a worldline given implicitly by $-t^2 + (x^1)^2 = \frac{1}{b^2}$ together with $x^2 = x^3 = 0$. Take the proper time s to be zero at $t = 0$. Show that the γ-factor of the particle is related to proper time s by $\gamma = \cosh bs$. e.o.e.

The four acceleration a^μ is defined by $a = \dot{u}$. Using (3.35), the four acceleration is related to the coordinate 3-acceleration $\vec{b} = \frac{d\vec{v}}{dt}$ by

$$a^\mu = \gamma \frac{d\gamma}{dt} \begin{pmatrix} 1 \\ \vec{v} \end{pmatrix} + \gamma^2 \begin{pmatrix} 0 \\ \frac{d\vec{v}}{dt} \end{pmatrix}\,,$$

or

$$a^\mu = \gamma^4 \vec{v}\vec{b} \begin{pmatrix} 1 \\ \vec{v} \end{pmatrix} + \gamma^2 \begin{pmatrix} 0 \\ \vec{b} \end{pmatrix} ,$$

The vector a^μ is spacelike. This follows from

$$\frac{d}{ds}(u,u) = 0 = 2(a,v) \tag{3.36}$$

The norm of a^μ is given by

$$(a,a) = \gamma^6 \vec{b}^2 (1 - |\vec{v}|^2 \sin^2 \alpha) , \tag{3.37}$$

where α is the angle between $\vec{v}$ and $\vec{b}$.

Exercise 30: Compute the four acceleration for the circular motion of Ex.19. e.o.e.

3.5 The relativistic rocket

We now consider a situation where a particle ('rocket') is accelerated by propelling a stream of particles through its exhaust. Consider first the Newtonian case. Suppose the rocket's mass is a given function $m(t)$ and $\vec{v}_1$ is the velocity of the exhaust gas in some inertial system. Then, by conservation of momentum, we have

$$\frac{d}{dt}(m\vec{v}) = -\dot{m}\,\vec{v}_1 \tag{3.38}$$

Now $\vec{v}_1 = \vec{w} - \vec{v}$, where $\vec{w}$ is the exhaust velocity in the ship's rest frame. Thus

$$m\dot{\vec{v}} = -\dot{m}\vec{w} \tag{3.39}$$

We assume the motion to be linear and $w = |\vec{w}|$ to be constant. Then (3.39) is readily integrated. When the ship moves from left to right, the exhaust is on the left and $v(0) = 0$, we get

$$\frac{m(t)}{m(0)} = e^{-\frac{v(t)}{w}} \tag{3.40}$$

Now to the relativistic case. The energy momentum carried away through the exhaust is given by a future-pointing causal vector J^μ. If u^μ is the four velocity of the rocket, the rate of change of its energy-momentum is given by minus J^μ. Thus

$$\frac{d}{ds}(mu^\mu) = -J^\mu , \tag{3.41}$$

where s is proper time of the spaceship. The current J^μ has to be proportional to the velocity four vector U^μ of the exhaust gas. Call this proportionality constant

λ. (Note we are not allowed here to use conservation of mass as in the Newtonian case.) Then (3.41) can be written as

$$\dot{m}u^\mu + m\dot{u}^\mu = -\lambda\, U^\mu\,, \tag{3.42}$$

where the overdot now denotes derivative w.r. to proper time and $\lambda > 0$ an as yet undetermined factor. Squaring (3.42) there results

$$-\dot{m}^2 + m^2(a,a) = -\lambda^2\,, \tag{3.43}$$

whereas taking the inner product of (3.42) with u yields

$$-\dot{m} = \lambda\,\gamma \tag{3.44}$$

with $\gamma = (1-w^2)^{-\frac{1}{2}}$, where $w = |\vec{w}|$ is the speed of the exhaust in the rocket's instantaneous rest frame. Eliminating λ from (3.43, 3.44) we obtain

$$\frac{\dot{m}}{m} = -\frac{\gamma(a,a)^{\frac{1}{2}}}{(\gamma^2-1)^{\frac{1}{2}}} = -\frac{(a,a)^{\frac{1}{2}}}{w} \tag{3.45}$$

Let again the motion be linear. Then, from (3.37),

$$(a,a)^{\frac{1}{2}} = |\vec{b}|(1-V^2)^{-\frac{3}{2}} = \frac{\dot{V}}{1-V^2}\,, \tag{3.46}$$

where the second equality in (3.46) uses that the exhaust is on the left. Assuming, again, that $V(0) = 0$ and w be constant, we see that

$$\frac{m(s)}{m(0)} = \left(\frac{1-\frac{V(s)}{c}}{1+\frac{V(s)}{c}}\right)^{\frac{c}{2w}} \tag{3.47}$$

By burning all its mass, the rocket can reach the speed of light.

Exercise 31: Reconsider the twin paradox. Suppose that the non-stationary twin has a photon rocket with an exhaust both on the front and back side. Suppose also she can burn 75% of her initial mass on the whole journey, and she can do so arbitrarily quickly. What is the best she can do w.r. to her ageing $\bar{s}$ relative to the ageing $\bar{t}$ of the stationary twin? Remark: at departure (arrival) the moving twin jumps on (off) the flying spaceship. So the engine is only used for decelerating and re-accelerating at the turning point. e.o.e.

Chapter 4

Relativistic field theory

4.1 Relativistic electrodynamics

Recall our discussion in Sect.(1.4) of the Lorentz invariance of the Maxwell equations, which in particular implies the following: when $(t, \mathbf{x})$ undergoes a Lorentz boost $(t, \mathbf{x}) \mapsto (\bar{t}, \bar{\mathbf{x}})$, there is an associated linear transformation $(\mathbf{E}, \mathbf{B}) \mapsto (\bar{\mathbf{E}}, \bar{\mathbf{B}})$, so that $(\bar{\mathbf{E}}, \bar{\mathbf{B}})(t(\bar{t}, \bar{\mathbf{x}}), \mathbf{x}(\bar{t}, \bar{\mathbf{x}}))$ solves the Maxwell equations provided $(\mathbf{E}, \mathbf{B})(t, \mathbf{x})$ does. (An analogous statement is true for rotations, where however the electric and magnetic field are not 'mixed' and the law relating components of $\mathbf{x}$ and $\bar{\mathbf{x}}$ and that relating $\mathbf{E}$ (resp. $\mathbf{B}$) to $\bar{\mathbf{E}}$ (resp. $\bar{\mathbf{B}}$) are the same.) The linear transformation, in the case of boosts, mapping $(\mathbf{E}, \mathbf{B})$ to $(\bar{\mathbf{E}}, \bar{\mathbf{B}})$ is essentially unexplained. Now, that we are in possession of Minkowski space, it is natural to try to represent the Maxwell equations as a system of partial differential equations for a tensor field on that spacetime, which are manifestly invariant under Lorentz (in fact: all Poincaré) transformations. Similary, it is tempting to try to write the law governing charged point particles - with the Lorentz force on the r.h. side - in relativistic language. To this we turn first.

The l.h. side has clearly to be of the form $\frac{d}{ds}(m\dot{z}^\mu)$, where $z(s)$ is the worldline of the charged particle parametrized by proper time and m its mass. When - as we assume - the mass is constant, the r.h. side F^μ has to be such that $(\dot{z}, F) = 0$, by virtue of (3.36). In a sense there is nothing wrong with that: we know from nonrelativistic theory that the laws of motion give three independent 2nd-order ODE's per particle. It is thus comforting to see that only three of the four equations

$$ma^\mu = F^\mu \tag{4.1}$$

are independent. But this also means that force fields $F^\mu(z, \dot{z})$ can not be invented out of the blue, as is possible in the nonrelativistic theory. On such rule could go as follows. Suppose, in the nonrelativistic version, a force law is given by $\mathbf{F} = -\nabla\Phi$, where Φ is a given function of $(t, \mathbf{x})$. Then a feasible relativistic version would be to define $F^\mu(z, \dot{z}) = \Pi^{\mu\nu}(\dot{z})(\partial_\nu\Phi)(z) = (\eta^{\mu\nu} + \dot{z}^\mu\dot{z}^\nu)(\partial_\nu\Phi)(z)$. In

fact, this is the force term which one would obtain in the simplest coupling of a point particle to a relativistic scalar field.
We now turn to electromagnetic theory and start with the familiar Lorentz force, namely

$$\mathbf{F} = e(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \tag{4.2}$$

We claim that the tensor equation

$$m\,\eta_{\mu\nu}\ddot{z}^\nu = eF_{\mu\nu}(z)\dot{z}^\nu \tag{4.3}$$

is a consistent relativistic version of $m\mathbf{b} = \mathbf{F}$ with $\mathbf{F}$ given by (4.2), when, in some inertial system, we set

$$F_{\mu\nu} = \begin{pmatrix} 0 & -E_1 & -E_2 & -E_3 \\ E_1 & 0 & B_3 & -B_2 \\ E_2 & -B_3 & 0 & B_1 \\ E_3 & B_2 & -B_1 & 0 \end{pmatrix}$$

In other words, we have that

$$F_{i0} = E_i \qquad F_{ij} = \epsilon_{ijk}B^k \tag{4.4}$$

Thus the i - component of (4.3) is given $\gamma(E_i + \epsilon_{ijk}v^jB^k)$. Furthermore, by the antisymmetry $F_{\mu\nu} = -F_{\nu\mu}$, we have identically $F_{\mu\nu}\dot{z}^\mu\dot{z}^\nu = 0$, so that only three equations are independent, as it has to be.
Let us next look at the expression $\partial^\nu F_{\mu\nu} = \eta^{\lambda\nu}\partial_\lambda F_{\mu\nu}$. We easily see that

$$\partial^\nu F_{0\nu} = -\partial^i E_i \qquad \partial^\nu F_{i\nu} = -\partial_t E_i + \epsilon_{ijk}\partial^j B^k \tag{4.5}$$

Thus, if we introduce the current four vector defined by

$$j^\mu = \begin{pmatrix} \rho \\ j^i \end{pmatrix}$$

the equations (1.59) can be rewritten in the manifestly Lorentz invariant form

$$\partial^\nu F_{\mu\nu} = 4\pi j_\mu \tag{4.6}$$

It is clear that the covector field j_μ can not be prescribed arbitrarily. Namely it follows from (4.6) that

$$\partial^\mu j_\mu = \partial_\mu j^\mu = 0\,, \tag{4.7}$$

which is the same as

$$\partial_t\rho + \nabla\mathbf{j} = 0 \tag{4.8}$$

which in turn is nothing but the differential version of charge conservation.
Next let us define the operation

$$\partial_{[\mu}F_{\nu\lambda]} = \frac{1}{3}\left(\partial_\mu F_{\nu\lambda} + \partial_\lambda F_{\mu\nu} + \partial_\nu F_{\lambda\mu}\right)\,. \tag{4.9}$$

Note that the left-hand side of (4.9) is totally antisymmetric, i.e. changes sign under permutation of an arbitrary pair of indices.

Exercise 32: Check that

$$3\,\partial_{[1}F_{23]} = \partial_i B^i \qquad 3\,\partial_{[0}F_{ij]} = \epsilon_{ijk}\dot{B}^k - \partial_j E_i + \partial_i E_j \tag{4.10}$$

and that (1.60) are equivalent to $\partial_{[\mu}F_{\nu\lambda]} = 0$. e.o.e.
Thus the full Maxwell equations are equivalent to the system (4.6) together with

$$\partial_{[\mu}F_{\nu\lambda]} = 0 \tag{4.11}$$

Finally we have to check that the behaviour of $F_{\mu\nu}$ under Lorentz transformations, namely $F_{\mu\nu} = L^\rho{}_\mu L^\sigma{}_\nu F'_{\rho\sigma}$ is consistent with what we found in Sect.(1.4). We defer this to the exercises below.
Exercise 33: Check, for a boost in the x^1 - direction, that the behaviour of $F_{\mu\nu}$ in terms of E_i and B_i is exactly that given in (1.63). e.o.e.
Exercise 34: Check, for the Lorentz transformation corresponding to a pure rotation, that the behaviour of $F_{\mu\nu}$ in terms of (E_i, B_i) is that of two Euclidean covectors under the Euclidean transformation given by a rotation. e.o.e.
Exercise 35: Suppose that $F_{\mu\nu}$ satisfies $F_{\mu\nu}F^{\mu\nu} = 0$, $F_{\mu\nu}F_{\rho\sigma} - F_{\mu\rho}F_{\nu\sigma} + F_{\mu\sigma}F_{\nu\rho} = 0$. Express these relations in terms of **E** and **B**. e.o.e.

4.2 The four potential A_μ

The Maxwell equations constitute a coupled system of 1st order partial differential equations. One would like to uncouple them, and this is one reason for introducing the so-called vector potential. There are also more fundamental reasons. Namely if one derives the Maxwell equations from an action principle, or if one couples the electromagnetic field to quantum objects - like a Dirac field - the vector potential becomes indispensable. In modern language Maxwell is - the simplest type of - a gauge theory.
Suppose there was a covector field A_μ, so that

$$F_{\mu\nu} = -2\,\partial_{[\mu}A_{\nu]} \tag{4.12}$$

Then (4.11) would be automatically satisfied. (Proof: Inserting (4.12) into the r.h. side of (4.11), one obtains 6 terms, out of which 2 triples cancel, due to $\partial_{[\mu}\partial_{\nu]}A_\lambda = 0$.) Interestingly the opposite is also true. Namely, given a field $F_{\mu\nu}$ satisfying (4.11), there is a field A_μ such that (4.12) is true.
We prove this statement by explicitly constructing a particular A_μ. Consider

$$A_\nu(x) = \int_0^1 F_{\nu\rho}(\lambda x)\,\lambda x^\rho d\lambda \tag{4.13}$$

then

$$\partial_{[\mu}A_{\nu]} = \int_0^1 [F_{\mu\nu}(\lambda x)\lambda + \partial_{[\mu}F_{\nu]\rho}(\lambda x)\lambda^2]x^\rho d\lambda \tag{4.14}$$

But, from (4.11),

$$\partial_{[\mu}F_{\nu]\rho} = -\frac{1}{2}\,\partial_\rho F_{\mu\nu} \tag{4.15}$$

Consequently the second term in (4.14) can be written as

$$-\frac{1}{2}\int_0^1 \left[\frac{d}{d\lambda}F_{\mu\nu}(\lambda x)\right]\lambda^2 d\lambda \tag{4.16}$$

Integrating (4.16) by parts, we obtain a term which cancels the first expression on the r.h. side of (4.14) and two boundary terms, one of which is zero. The remaining boundary term is exactly $-\frac{1}{2}\,F_{\mu\nu}(x)$. Thus (4.13) solves (4.12).
A vector potential A_μ for $F_{\mu\nu}$ is not unique. Namely, given A_μ, the covector field

$$\bar{A}_\mu = A_\mu + \partial_\mu \Lambda \tag{4.17}$$

also solves (4.12) for an arbitrary function $\Lambda(x)$. The map sending A_μ to $\bar{A}_\mu$ is called a gauge transformation. Next turn to the remaining Maxwell equations (4.6). They yield

$$\Box A_\mu - \partial_\mu\partial^\nu A_\nu = 4\pi j_\mu \tag{4.18}$$

(4.18) would be a system of (uncoupled) wave equations were it not for the presence of the 2nd term. Now the gauge dependence of A_μ comes as a blessing: we can try to find a gauge transformation, so that $\bar{A}_\mu$ satisfies $\partial^\mu \bar{A}_\mu = 0$. This is possible: just solve the equation $\Box\Lambda = -\partial^\mu A_\mu$ and set $\bar{A}_\mu = A_\mu + \partial_\mu\Lambda$. This done, the Maxwell equations take the beautiful form

$$\Box A_\mu = 4\pi j_\mu \tag{4.19}$$

Exercise 36: Show that $\partial_\mu \int_0^1 \Lambda_\nu(\lambda x)x^\nu d\lambda = \Lambda_\mu(x)$, provided that $\partial_{[\mu}\Lambda_{\nu]} = 0$. e.o.e.

We end this section by counting degrees of freedom of the electromagnetic field at a point. This is done as follows. We make the ansatz

$$A_\mu(x) = a_\mu\, e^{i\,k_\nu x^\nu} \tag{4.20}$$

where k_μ and a_μ are constant covectors and assume the gauge $\partial^\mu A_\mu = 0$, which is the Lorenz gauge (often wrongly called 'Lorentz gauge'). It follows that

$$(a,k) = 0 \qquad \text{and} \quad (k,k) = 0 \tag{4.21}$$

Thus k is a null (co-) vector, which - in the case of the wave equation - we already know from Sect.2.7. Suppose we make yet another gauge transformation of the for $A_\mu \mapsto A_\mu + \lambda_\mu \, e^{i\,k_\nu x^\nu}$. When $\lambda_\mu \sim k_\mu$, this preserves the Lorenz gauge. Thus, given the lightlike covector k_μ, the degrees of freedom are the set of covectors a orthogonal to the k, modulo the transformation $a_\mu \mapsto a_\mu + l\, k_\mu$, that-is-to-say: 2. We can describe the allowed c's more concretely by picking another null vector l, linearly independent from k, normalized say by $(k,l) = -2$ and requiring that a satisfy $(a,l) = 0$ in addition to $(a,k) = 0$. For example k correspond to a light ray along the positive x^1 - axis and l one along the negative x^1 - axis, in which case the allowed a's are of the form $a = (0, 0, a^2, a^2)$.

Exercise 37: Discuss the structure of $F_{\mu\nu}$ corresponding to A_μ as in (4.20,4.21) in the context of Exercise 35. e.o.e.
Exercise 38: Let $\mathbf{EB}$ be zero and $\mathbf{E}^2 - \mathbf{B}^2 \neq 0$. Show that there is a Lorentzian frame so that either $\mathbf{E}$ or $\mathbf{B}$ vanishes. (Hint: Take $\mathbf{v}$ proportional to $\mathbf{E} \times \mathbf{B}$.) e.o.e.

4.3 The Lienard-Wiechert fields

Let $z(s)$ be a timelike future-pointing curve and x an arbitrary event. Then $\mathcal{C}_+(x)$ intersects the curve in the unique point $z(s_+)$ and $\mathcal{C}_-(x)$ intersects the curve in $z(s_-)$. The values $s_\pm$ are called the retarded (advanced) times. They are as functions of x implicitly given by the equation

$$(x - z(s_\pm), x - z(s_\pm)) = 0 \tag{4.22}$$

Now define

$$A^\mu_\pm(x) = \mp e \, \frac{v^\mu_\pm}{(v_\pm, x - z_\pm)} \tag{4.23}$$

where $z_\pm = z(s_\pm)$, $v_\pm = v(s_\pm)$ and x lies outside of $z(s)$. The fields A_+ (resp. A_-) are called the advanced (retarded) Lienard-Wiechert potentials for a charge moving along $z(s)$. We claim that $A_\pm$ solves the equations (4.19) for $j_\mu = 0$ together with the Lorenz gauge. We check this latter fact and leave the verification of $\Box A_\pm = 0$ as an exercise. By implicit differentiation of $(x-z, x-z) = 0$ we infer that

$$\partial^\mu s_\pm = \frac{(x - z_\pm)^\mu}{(v_\pm, x - z_\pm)} \tag{4.24}$$

Thus

$$\begin{aligned} \mp \frac{1}{e}\, \partial^\nu A^\mu_\pm = \frac{\dot v^\mu_\pm (x - z_\pm)^\nu}{(v_\pm, x - z_\pm)^2} - \frac{v^\mu_\pm (\dot v_\pm, x - z_\pm)(x - z_\pm)^\nu}{(v_\pm, x - z_\pm)^3} - \\ - \frac{v^\mu_\pm v^\nu_\pm}{(v_\pm, x - z_\pm)^2} + \frac{v^\mu_\pm (v_\pm, v_\pm)(x - z_\pm)^\nu}{(v_\pm, x - z_\pm)^3} \,, \end{aligned} \tag{4.25}$$

and from this it is easy to infer that $\eta_{\mu\nu}\,\partial^\nu A^\mu_\pm = 0$.

Exercise 39: (Coulomb field) Take the case of a charge which, in some inertial system, is at rest at the spatial point $\mathbf{z}_0$. Thus $z^\mu(s) = (s, \mathbf{z}_0)$. Calculate $A_\pm$ and show that

$$\Box A^\mu_\pm(x) = 4\pi e\,\delta^{(3)}(\mathbf{x} - \mathbf{z}_0)\,\dot z^\mu \tag{4.26}$$

We state without proof that the result of this exercise can be generalized to arbitrary motions by saying that, in the sense of distributions, there holds

$$\Box A^\mu_\pm(x) = 4\pi e \int_{-\infty}^{\infty} \delta^{(4)}(x - z(s))\,\dot z^\mu(s)\,ds \tag{4.27}$$

Note that the r.h. side of (4.27) is the same as $4\pi e\,\delta^{(3)}(\mathbf{x} - \mathbf{z}(t))\frac{dz^\mu(t)}{dt}$.

4.4 The energy momentum tensor

There is an energy momentum tensor - also called 'stress-energy tensor' for every relativistic field theory. It is conceptually not very easy to grasp. Its true meaning unfolds only in the context of general relativity. We now introduce this concept in the case of the electromagnetic field. Let $F_{\mu\nu}$ satisfy the Maxwell equations for some source j_μ. Then define a symmetric tensor field $T_{\mu\nu}$ by

$$T_{\mu\nu} = \frac{1}{4\pi}\,[F_{\mu\rho}F_\nu{}^\rho - \frac{1}{4}\,\eta_{\mu\nu}\,F_{\rho\sigma}F^{\rho\sigma}] \tag{4.28}$$

In addition to being symmetric, $T_{\mu\nu}$ is trace-free, i.e.

$$\eta^{\mu\nu}T_{\mu\nu} = 0 \tag{4.29}$$

Computing $T_{\mu\nu}$ in some inertial frame we find, using (4.4), that

$$4\pi\,T_{00} = \frac{1}{2}(\mathbf{E}^2 + \mathbf{B}^2) \tag{4.30a}$$

$$4\pi\,T_{0i} = -\,(\mathbf{E}\times\mathbf{B})_i = -J_i \tag{4.30b}$$

$$4\pi\,T_{ij} = \frac{1}{2}\,\delta_{ij}(\mathbf{E}^2 + \mathbf{B}^2) - E_iE_j - B_iB_j \tag{4.30c}$$

The covector J_i on the r.h. side of (4.30b) is the well-known Poynting vector. Let us compute the divergence of $T_{\mu\nu}$, i.e. $\partial^\nu T_{\mu\nu}$. Using (4.6) we find

$$\partial^\nu T_{\mu\nu} = -F_{\mu\nu}j^\nu + \frac{1}{4\pi}\,[F^{\nu\rho}\partial_\nu F_{\mu\rho} - \frac{1}{2}F^{\rho\sigma}\partial_\mu F_{\rho\sigma}] \tag{4.31}$$

But apart from a factor $-\frac{1}{8\pi}$ the second term in (4.31) is the same as

$$F^{\rho\sigma}(2\,\partial_\sigma F_{\mu\rho} + \partial_\mu F_{\rho\sigma}) = F^{\rho\sigma}(\partial_\sigma F_{\mu\rho} - \partial_\rho F_{\mu\sigma} + \partial_\mu F_{\rho\sigma}) = 0\,, \tag{4.32}$$

where we have used (4.11) in the last step. Thus

$$\partial^\nu T_{\mu\nu} = -F_{\mu\nu} j^\nu \tag{4.33}$$

To understand the nature of this equation, we have to keep in mind that a conservation law in continuum mechanics has the form of a continuity equation, or, in relativistic language, has the form of a 'conserved current', i.e. a four vector which is divergence free. We have just encountered the example of charge conservation. We encounter further examples here in the form of energy conservation and momentum conservation for the electromagnetic field. To each of these quantities there should correspond a conserved current. Suppose for simplicity that $j^\mu = 0$. Then (4.33) can be thought of a supplying four such conserved currents, one for each index μ, which exactly correspond to these conservation laws. The corresponding balance laws, using (4.33) and the Gauss theorem, are as follows

$$\frac{d}{dt}\int_V T^{00} d^3x = -\int_{\partial V} T^{0i}\, dS_i \tag{4.34a}$$

$$\frac{d}{dt}\int_V T^{0i} d^3x = -\int_{\partial V} T^{ij} dS_j\ , \tag{4.34b}$$

where V is a time-independent three volume with boundary ∂V. The vector field $\mathcal{P}^\mu$ with components (T^{00}, T^{0i}) can be thought of as the conserved current corresponding to the energy, the collection of vector fields (T^{0j}, T^{ij}) for each j be thought of as conserved currents corresponding to the components of momentum. Thus, from standard continuum mechanics, T^{ij} plays the role of stress tensor. The former current can also be described as follows. Let u^μ be the constant vector field with $(u, u) = -1$ corresponding to some inertial observer. Then $\mathcal{P}^\mu(u) = -T^\mu{}_\nu u^\nu$ is the 'energy current' for the observer u. The vector $\mathcal{P}^\mu(u)$ is future-pointing timelike or null.

$$-(T^0{}_0)^2 + T^0{}_i T^{0i} = -\frac{1}{4}(\mathbf{E}^2 + \mathbf{B}^2)^2 + (\mathbf{E}\times\mathbf{B})^2 = -\frac{1}{4}(\mathbf{E}^2 - \mathbf{B}^2)^2 - (\mathbf{E}\mathbf{B})^2 \tag{4.35}$$

The above stress tensor (called 'Maxwell stress tensor') can be more invariantly described as the four dimensional tensor $t^{\mu\nu}(u) = \Pi^\mu{}_\rho(u)\, \Pi^\nu{}_\sigma(u)\, T^{\rho\sigma}$.

Exercise 40: Show that

$$T_{\mu\rho} T_\nu{}^\rho = \frac{1}{4}\, \eta_{\mu\nu} T_{\rho\sigma} T^{\rho\sigma} \tag{4.36}$$

4.5 Energy momentum tensor of dust

Consider first the case of 'dust': this is to be thought of as a continuous collection of particles, which neither interact nor collide. The only available quantities are first the vector field v^μ (with $(v,v)=-1$) of the 4-velocity of these particles: this describes the 'flow' $\phi^\mu(s\,;x)$, mathematically given by the system of ODE's

$$\frac{d\phi^\mu}{ds} = v^\mu(\phi) \tag{4.37}$$

with the initial condition $\phi^\mu(0;x) = x^\mu$. Second, there is ρ, the mass density of the dust, i.e. the energy density of the dust in in its own rest frame. We require $\rho > 0$. The simplest ansatz is given by

$$T^{\mu\nu} = \rho\, v^\mu v^\nu\,. \tag{4.38}$$

In accordance with our experience with electrodynamics we identify $\mathcal{E}(u) = T_{\mu\nu}u^\mu u^\nu$ with the energy density of the dust in the rest frame of the observer given by u^μ (with $(u.u) = -1$). Thus

$$\mathcal{E}(u) = \rho\,\gamma^2\,, \tag{4.39}$$

where γ refers to the relative velocity between u and v. Why does γ^2 appear as factor? The answer is this: one factor of γ comes from the velocity dependence of the energy of a single particle, as in (3.7). The other factor of γ comes from the (inverse) Lorentz contraction of volumes in the direction of relative motion (recall that energy density is energy/volume). So (4.39) is in fact in accordance with expectation. We now impose the law $\partial_\nu T^{\mu\nu} = 0$. Since

$$\partial_\nu T^{\mu\nu} = v^\mu \partial_\nu(\rho v^\nu) + \rho v^\nu \partial_\nu v^\mu\,, \tag{4.40}$$

we find, by scalar multiplication of (4.40) with u_μ, that

$$\partial_\mu(\rho\, v^\mu) = 0\,. \tag{4.41}$$

Inserting (4.41) back into (4.40) we finally obtain

$$v^\nu \partial_\nu v^\mu = 0 \tag{4.42}$$

The meaning of (4.41) is that of conservation of mass. While this is not a fundamental law of relativistic physics, it does hold in the very special case of dust.
To wee what equation (4.42) means recall that the trajectory of a single particle is given by a solution $x^\mu = z^\mu$ to (4.37) with some fixed initial data, e.g. $z^\mu(s) = \phi^\mu(s;x_0)$. We now compute $\ddot{z}^\mu(s)$. We find

$$\ddot{z}^\mu(x) = \frac{d}{ds}\, v^\mu(z(s)) = v^\nu \partial_\nu v^\mu\,. \tag{4.43}$$

Thus (4.42) says nothing but that individual particles follow straight lines.

Exercise 41: Let x^μ be the 4 dimensional position vector field, u^μ a constant vector field with $(u, u) = -1$ and restrict to points x away from the timelike line given by $(x, x) + (x, u)^2 = 0$. Define a vector field v^μ by

$$v = \sqrt{1 + W^2}\, u - W\, \frac{x + (x, u)\, u}{\sqrt{(x, x) + (x, u)^2}}\ , \tag{4.44}$$

where W is a constant. Discuss the physical meaning of the 'flow' described by v^μ. (Hint: take a Lorentzian frame adapted to u^μ). Prove that $v^\nu \partial_\nu v^\mu = 0$. What happens when W is no longer constant, but still $u^\mu \partial_\mu W = 0$ and $[x + (x, u)u]^\mu\, \partial_\mu W = 0$? e.o.e.

4.6 Energy momentum tensor of perfect fluid matter

The word 'perfect' for fluids (or solids) refers to the absence of irreversible phenomena like heat transfer or friction. It is then plausible to assume that, in analogy to free particles, the energy density $\mathcal{E}(w) = T_{\mu\nu} w^\mu w^\nu$ assumes a minimum for some timelike vector field v^μ - the tangent vector to the fluid flow (with $(v, v) = -1$). Hence taking an arbitrary 1 parameter family of timelike vector fields $w(\epsilon)$ with $w^\mu(0) = v^\mu$, so that $\frac{d}{d\epsilon}\, \mathcal{E}(w(\epsilon))|_{\epsilon=0} = 0$. Now calculating this latter expression gives

$$2\, T_{\mu\nu} X^\mu v^\nu = 0\ , \tag{4.45}$$

where $X^\mu = \frac{d}{d\epsilon} w^\mu|_{\epsilon=0} = 0$. But differentiating $(w(\epsilon), w(\epsilon)) = -1$ w.r. to ϵ and setting ϵ equal to zero, we infer that $(X, v) = 0$, so X is spacelike. But since X can, by an appropriate choice of sequence $w(\epsilon)$, be taken to be an arbitrary vector orthogonal to v, the covector $T_{\mu\nu} v^\mu$ has to be proportional to v_μ. Thus there is a function ρ, so that $T^\mu{}_\nu v^\nu = \rho v^\mu$. The quantity ρ is of course nothing but $\mathcal{E}(v)$ and will thus be assumed to be non-negative. Now consider the tensor field $t^{\mu\nu}$ given by

$$t^{\mu\nu} = T^{\mu\nu} - \rho v^\mu v^\nu \tag{4.46}$$

From the above it has the property of being 'purely spatial' in the sense that $t^{\mu\nu} v_\nu = 0$ and is thus the stress tensor corresponding to $T^{\mu\nu}$ in the rest system of the fluid. Next recall that fluids, as opposed to elastic solids, have the property that the force exerted on any infinitesimal cube of material is the same for all faces of this cube. This then implies that the stress tensor of a fluid is 'isotropic' which in turn means that it is proportional to the Kronecker delta (all eigenvalues are equal). The factor of proportionality is nothing but the pressure p. Let us, more concretely, pick an event x_0 and a Lorentzian frame at this event, for which $v^\mu = (1, 0, 0, 0)$. Then our requirement takes the form that there is a function

p, so that, in this 'comoving' frame, $T^{ij} = p\,\delta^{ij}$. Next observe that, still in the comoving frame, $t^{ij} = T^{ij}$ and $t^{0i}\,, t^{ij}$ are both zero. Note finally that the field $\Pi^{\mu\nu}(v) = \eta^{\mu\nu} + v^\mu v^\nu$ has the property that $\Pi^{0i}(v)\,, \Pi^{00}(v)$ are both zero in the frame comoving with v and that $\Pi^{ij}(v) = \delta^{ij}$ in this frame. Thus our requirement takes the final form

$$T^{\mu\nu} - \rho\, v^\mu v^\nu = p\,(\eta^{\mu\nu} + v^\mu v^\nu) \tag{4.47}$$

or

$$T^{\mu\nu} = (\rho + p)\, v^\mu v^\nu + p\,\eta^{\mu\nu} \tag{4.48}$$

Again we assume that the equations of motion are given by the law $\partial_\nu T^{\mu\nu} = 0$. We write this as

$$\partial_\nu[(\rho + p)v^\nu]\, v^\mu + (\rho + p)\, v^\nu \partial_\nu v^\mu + \partial^\mu p = 0 \tag{4.49}$$

Contracting (4.49) with v_μ, there results

$$\partial_\mu(\rho v^\mu) + p\,\partial_\mu v^\mu = 0 \tag{4.50}$$

Inserting (4.50) back into (4.49), we finally infer

$$(\rho + p)\, v^\nu \partial_\nu v_\mu + (\delta_\mu{}^\nu + v_\mu v^\nu)\,\partial_\nu p = 0\ , \tag{4.51}$$

which is the relativistic form of the Euler equation of nonrelativistic hydrodynamics. To compare (4.50,4.51) with their nonrelativistic versions observe that ρ has the same physical dimension as $\frac{p}{c^2}$. Thus we end up with

$$\left(\rho + \frac{p}{c^2}\right)\, v^\nu \partial_\nu v_\mu + (\delta_\mu{}^\nu + v_\mu v^\nu)\,\partial_\nu p = 0 \tag{4.52}$$

together with

$$\partial_\mu(\rho v^\mu) + \frac{p}{c^2}\,\partial_\mu v^\mu = 0\ , \tag{4.53}$$

provided that v^μ has the physical dimension of 'velocity'. The presence of the p-term in (4.50,4.53) shows that the mass equivalent of the energy density of the fluid is not generally conserved, except in the very special case of a flow with $\partial_\mu v^\mu = 0$ (which can be shown to correspond to incompressible motion of the fluid).

The relations (4.52,4.53) furnish 4 equations for the 5 unknowns (ρ, p, v): Eq.(4.52) is orthogonal to v, whence these are only 3 equations. And v^μ satisfies the constraint $(v, v) = -1$, so these are only 3 independent quantities. Thus our system of equations is not yet fully determined. The missing piece of information is usually provided by thermodynamic (or statistical mechanical) information, for instance in the form that p be a given function of ρ. For example if the fluid is composed of 'pure radiation', i.e. zero rest-mass particles on a microscopic level, then statistical mechanics shows that the energy density is related to the pressure through the equation $\rho = \frac{p}{3}$. Note that in this case, although microscopically particles are moving at the speed of light, the motion-at-large (averaged motion) of the system is still given by v^μ, i.e. subluminal.

Chapter 5

Basics of general relativity

5.1 The equivalence principle

We start by summarizing the essence of special relativity.
First recall that Minkowski spacetime is defined as a 4-dimensional vector (or affine) space endowed with a metric $\eta_{\mu\nu}$ with signature diag$(-,+,+,+,)$. Next there is the mathematical fact that the Maxwell theory allows a formulation which is 'natural' in terms of that geometrical structure. In particular the laws of this theory are invariant under isometries of Minkowski space, that-is-to-say Poincaré transformations, which contain translations, spatial rotations and Lorentz boosts. The latter of which reduce to Galilei boosts in the limit where $c \to \infty$ and can thus be interpreted as transitions between different inertial systems. Moreover, on the experimental side there were never found any effects of a preferred ('ether') inertial system for electromagnetic phenomena. Based on these pieces of evidence one makes the decision that the Poincaré symmetry is the 'right' symmetry on which both the theories of electrodynamics and mechanics - and in fact: all of physics - should be based. While this program eventually met with enormous success in the realm of high energy physics, it was realized much earlier that the gravitational field is an exception. The heuristic reason for this lies in the famous equivalence principle. This is based on the following sequence of considerations: First recall the equations governing a (point) particle falling freely in (i.e. 'subject to no other forces than') a gravitational field, namely

$$m_i\, \ddot{\mathbf{x}} = -\, m_g \nabla\Phi(\mathbf{x}) \tag{5.1}$$

Here Φ denotes the Newtonian gravitational potential, and m_i, m_s are respectively the inertial and gravitational mass. The former concept is the kind of mass referred to in Newton's second law ('force equals mass times acceleration'), whereas the second concept means 'mass as a constant measuring the strength of the coupling of the particle to the gravitational field'. Experimentally it is known with an accuracy of 10^{-12}, that $m_i = m_g$. Note that the law of motion for freely falling particles has the form (5.1) in some inertial system. When we transform to an accelerated system, given say by $\mathbf{y} = \mathbf{x} - \frac{1}{2}\mathbf{g}t^2$, this equation changes to

$$\ddot{\mathbf{y}} = -\nabla\Phi(\mathbf{y} + \frac{1}{2}\mathbf{g}t^2) - \mathbf{g} \tag{5.2}$$

Suppose we confine attention to a small enough region in $(t, \mathbf{y})$ - space, so that $\nabla\Phi$ can be assumed to be approximately constant. It then follows that the effect of gravity can be compensated by a suitable choice of $\mathbf{g}$. Conversely, for a particle subject to no forces whatsoever, the effect of gravity can be simulated by a suitable choice of $\mathbf{g}$. These statements are called the equivalence principle. Next recall that free particles in special relativity are described by a worldline $x^\mu(\lambda)$ with timelike, future-pointing tangent $v^\mu(\lambda) = \dot{x}^\mu(\lambda)$ satisfying (4.43)

$$\ddot{x}^\mu = \dot{x}^\nu \partial_\nu \dot{x}^\mu = 0 \tag{5.3}$$

In arbitrary coordinates this can, following Appendix A, be written as

$$\ddot{x}^\mu + \Gamma^\mu_{\nu\rho} \dot{x}^\nu \dot{x}^\rho = 0 \; . \tag{5.4}$$

Now comes the idea for a mathematical formulation of the law governing the motion of a relativistic particle falling freely in a gravitational field: This motion is still described by equation (5.3), but where the Christoffel symbols now refer to a spacetime metric which is not the (necessarily) the Minkowski metric any longer. It is such a curved spacetime in which the effect of the gravitational field is encoded. Then the equality of inertial and gravitational mass is no longer an additional postulate required by experiment, but an automatic consequence: only the sum of the second derivative and the Christoffel term in (5.3) have a coordinate independent meaning (via the concept of covariant derivative with respect to some - not necessarily flat - metric). And it is this Christoffel term, which plays the role of the force acting on the particle. We have to spend the next sections making the above statements precise.

5.2 Curved spacetimes

First recall the notion of a tensor field of valence (0,2). This is an array of functions $g_{\mu\nu}(x)$, so that, for any coordinate transformation $\bar{x}(x)$,

$$\bar{g}_{\mu\nu}(\bar{x}) = \frac{\partial \bar{x}^\rho}{\partial x^\mu}(x(\bar{x})) \frac{\partial \bar{x}^\sigma}{\partial x^\nu}(x(\bar{x})) g_{\rho\sigma}(x(\bar{x})) \tag{5.5}$$

This tensor field is called symmetric, when $g_{\mu\nu} = g_{\nu\mu}$. It is easy to check that this holds in any coordinate system when it holds in one. Now consider a coordinate change, so that $\bar{x}(x_0) = x_0$ for some point x_0. Then

$$\bar{g}_{\mu\nu}(x_0) = A^\rho{}_\mu A^\sigma{}_\nu \, g_{\rho\sigma}(x_0) \tag{5.6}$$

where $A^\rho{}_\mu = \frac{\partial \bar{x}^\rho}{\partial x^\mu}(x_0)$. Now, it is a basic fact from linear algebra that, for any quadratic form like $g_{\mu\nu}(x_0)$, there is a nonsingular matrix A, so that $\bar{g} = A^T g A$ is diagonal with $0, 1, -1$ as the only elements. The tensor field $g_{\mu\nu}(x)$ is called a Lorentz metric at x_0, when there is a matrix A, so that $A^T g(x_0) A$ is $\text{diag}(-1, 1, 1, 1)$. The tensor field $g_{\mu\nu}(x)$ is called a Lorentz or spacetime metric when it is Lorentz at all points. A space equipped with a spacetime metric is called a spacetime. The most important example of a spacetime is of course Minkowski space, the arena of special relativity, where the metric is given, in some coordinate system, by $g_{\mu\nu}(x) = \eta_{\mu\nu} = \text{diag}(-1, 1, 1, 1)$. A spacetime metric $g_{\mu\nu}(x)$ is called flat, when it is Minkowski in a suitable coordinate system. Given a metric in some arbitrary coordinates, it may be hard to recognize, if the metric is flat. We will give a criterion later.
Suppose $g_{\mu\nu}(x)$ is a Lorentz metric. It follows that $g_{\mu\nu}$ viewed as a linear map from vector to covectors, is invertible (e.g. its determinant is everywhere non-zero), and consequently there exists the inverse map $g^{\mu\nu}(x)$, sending covectors into vectors, which is again symmetric in the sense that $g^{\mu\nu} = g^{\nu\mu}$. Thus there is a concept of index-lowering and index-raising like in Minkowski space.
Next the concepts of timelike, spacelike, null vectors (and covectors) carry over literally from special relativity, likewise that of a timelike, null, etc. curve. In addition these concepts are well-defined, i.e. do not depend on the choice of coordinate system.
We now come to the notion of covariant derivative of tensor fields. So far the covariant derivative was just the standard partial derivative in a coordinate system where the metric (either the Euclidean one for $n = 3$ or the Minkowski one for $n = 4$) had the standard form. We generalize this to a situation where the metric is not necessarily flat. To find this generalization, it is simplest to require, for the sought-after covariant derivative, a set of rules generalizing those which are valid for the flat covariant derivative, and then show that covariant derivative exists and is unique. Here are these rules:
(o) Let s and t be tensor fields of the same type. Then

$$\nabla_\mu(s + t) = \nabla_\mu s + \nabla_\mu t \tag{5.7}$$

(i) Let f be a scalar field. Then $\nabla_\mu f$, the covariant derivative of f, is given by

$$\nabla_\mu f = \partial_\mu f \; . \tag{5.8}$$

Note that, by Sect.(1.2), the r.h. side of (5.8), is a covector field.
(ii) The covariant derivative satisfies the Leibniz rule with respect to arbitrary tensor products, e.g.

$$\nabla_\mu(\omega_\nu \, t^{\rho\sigma}) = (\nabla_\mu \omega_\nu) \, t^{\rho\sigma} + \omega_\nu \nabla_\mu t^{\rho\sigma} \tag{5.9}$$

(iii) 'Vanishing torsion':

$$\nabla_\mu \nabla_\nu f = \nabla_\nu \nabla_\mu f \tag{5.10}$$

(iv) 'Covariant derivation commutes with contraction':

$$\nabla_\mu(\omega_\nu v^\nu) = (\nabla_\mu \omega_\nu) v^\nu + \omega_\nu \nabla_\mu v^\nu \tag{5.11}$$

Note that, by (5.8), $\nabla_\mu(\omega_\nu v^\nu) = (\partial_\mu \omega_\nu) v^\nu + \omega_\nu \partial_\mu v^\nu$ holds trivially, whereas (5.11) does not. Finally we require
(v) 'metricity':

$$\nabla_\mu g_{\nu\rho} = 0 \tag{5.12}$$

A covariant derivative satisfying the postulates (o) to (v) exists. Namely define

$$\nabla_\mu t^{\nu..\tau}{}_{\rho..\sigma} = \partial_\mu t^{\nu..\tau}{}_{\rho..\sigma} + \Gamma^\nu_{\mu\chi} t^{\chi..\tau}{}_{\tau..\sigma} + ..\Gamma^\tau_{\mu\chi} t^{\nu..\chi}{}_{\tau..\sigma} - ..\Gamma^\chi_{\mu\rho} t^{\nu..\tau}{}_{\chi..\sigma} - ..\Gamma^\chi_{\mu\sigma} t^{\nu..\tau}{}_{\rho..\chi} \tag{5.13}$$

where $\Gamma^\mu_{\nu\rho}$ is given by (6.17), namely

$$\Gamma^\mu_{\nu\rho} = \frac{1}{2} g^{\mu\sigma} (\partial_\nu g_{\sigma\rho} + \partial_\rho g_{\sigma\nu} - \partial_\sigma g_{\nu\rho}) \tag{5.14}$$

Exercise 42: Check that (5.13) does indeed satisfy (o) to (v). e.o.e.

Thus a covariant derivative satisfying our axioms exists. In fact, the chosen covariant derivative is also unique. First of all, once the action of ∇ on vectors and covectors is fixed, it is clearly fixed for tensors of arbitrary rank: this is due to the Leibniz rule (ii) and the fact that all tensors are linear combinations of tensor products of vectors and covectors. Secondly, the action of ∇ on covectors is fixed, once it is fixed on vectors, due to the axiom (iv). It remains to show that $\nabla_\mu v^\nu$ has to have the form

$$\nabla_\mu v^\nu = \partial_\mu v^\nu + \Gamma^\mu_{\nu\rho} v^\rho \ , \tag{5.15}$$

where $\Gamma^\mu_{\nu\rho}$ is given by (5.14). Consider the quantity $\Delta_\mu v^\nu$ given by

$$\Delta_\mu v^\nu = (\nabla_\mu - \partial_\mu) v^\nu \tag{5.16}$$

From axioms (i,ii) it follows that

$$\Delta_\mu (\phi\, v^\nu) = \phi\, \Delta_\mu v^\nu \tag{5.17}$$

for every scalar field ϕ. Let $e^\nu_{(\mu)}$ be a collection of basis vector fields and define

$$C^\mu_{\nu\rho} = \Delta_\nu\, e^\mu_{(\rho)} \ . \tag{5.18}$$

Then, using (5.17), it follows that

$$\Delta_\mu v^\nu = C^\nu_{\mu\lambda} v^\lambda \tag{5.19}$$

for all vector fields v^μ. It follows from this and our previous remarks that for example

$$\nabla_\mu \omega_\nu = \partial_\mu \omega_\nu - C^\rho_{\mu\nu}\omega_\rho \tag{5.20}$$

Using axiom (iii), it follows that

$$C^\rho_{\mu\nu}\partial_\rho \phi = C^\rho_{\nu\mu}\partial_\rho \phi \tag{5.21}$$

for all scalar fields ϕ. Choosing for example ϕ to be a linear function in the given coordinates, it becomes clear that this requires that

$$C^\rho_{\mu\nu} = C^\rho_{\nu\mu} \tag{5.22}$$

It remains to use metricity, i.e. that

$$\nabla_\mu g_{\nu\rho} \equiv \partial_\mu g_{\nu\rho} - C^\sigma_{\mu\nu} g_{\sigma\rho} - C^\sigma_{\mu\rho} g_{\nu\sigma} = 0 \tag{5.23}$$

Now using an argument completely equivalent to that after (6.14), it follows that the unique solution of (5.23) is given by $C^\sigma_{\mu\nu} = \Gamma^\sigma_{\mu\nu}$. This completes the uniqueness proof.
We remark (again) that $\Gamma^\sigma_{\mu\nu}$ do not form components of a tensor field.

Exercise 43: Show that

$$\frac{\partial x^\nu}{\partial \bar{x}^\rho}\frac{\partial x^\lambda}{\partial \bar{x}^\sigma}\frac{\partial^2 \bar{x}^\mu}{\partial x^\nu \partial x^\lambda} = -\frac{\partial \bar{x}^\mu}{\partial x^\tau}\frac{\partial^2 x^\tau}{\partial \bar{x}^\rho \partial \bar{x}^\sigma} \tag{5.24}$$

Exercise 44: Show that $\Gamma^\mu_{\nu\lambda}$ transforms as (compare (6.2))

$$\bar{\Gamma}^\mu_{\nu\lambda} = \frac{\partial \bar{x}^\mu}{\partial x^\tau}\frac{\partial x^\rho}{\partial \bar{x}^\nu}\frac{\partial x^\sigma}{\partial \bar{x}^\lambda}\Gamma^\tau_{\rho\sigma} - \frac{\partial x^\rho}{\partial \bar{x}^\nu}\frac{\partial x^\sigma}{\partial \bar{x}^\lambda}\frac{\partial^2 \bar{x}^\mu}{\partial x^\rho \partial x^\sigma} \tag{5.25}$$

The following definition characterizes Special Relativity as a special case of the present scenario. A (curved) spacetime is called flat, when there exist coordinates so that $\Gamma^\mu_{\nu\lambda} = 0$. Then, by the identity (compare (6.16))

$$\partial_\mu g_{\nu\rho} - \Gamma^\sigma_{\mu\nu} g_{\sigma\rho} - \Gamma^\sigma_{\mu\rho} g_{\nu\sigma} = 0 \ , \tag{5.26}$$

it follows that the metric is constant in the given coordinate system x^μ. By a linear transformation of x^μ we can arrange that $g_{\mu\nu} = \eta_{\mu\nu}$. Consequently the metric is the Minkowski metric.

Exercise 45: Let ξ^μ and η^μ be vector fields. Show that $[\xi, \eta]^\mu = \xi^\nu \partial_\nu \eta^\mu - \eta^\nu \partial_\nu \xi^\mu$ is also a vector field (i.e. transforms like such). e.o.e.
Exercise 46: Prove the 'Jacobi identity' identity for vector fields, i.e.

$$[[u, v], w] + [[w, u], v] + [[v, w], u] = 0 \tag{5.27}$$

Exercise 47: Let $\xi = \xi^\mu \partial_\mu$ be a vector field and (t, x^i) coordinates such that $\xi^\mu \partial_\mu = \partial_t$. Now consider coordinate transformations of the form

$$(t, x^i) \mapsto (\bar{t} = t - F(x), \bar{x}^i = f^i(x^j)) . \tag{5.28}$$

Show that we still have $\xi = \partial_{\bar{t}}$. e.o.e.

5.3 Again: equivalence principle

Let $x^\mu(\lambda)$ be a curve in spacetime. This curve is called a geodesic, when

$$v^\nu \nabla_\nu v^\mu \equiv \ddot{x}^\mu + \Gamma^\mu_{\nu\rho} \dot{x}^\nu \dot{x}^\rho = 0 \ , \tag{5.29}$$

where $v^\mu = \dot{x}^\mu$ is the tangent vector. Clearly in flat spacetime geodesics are straight lines. Geodesics in curved spacetime have a number of properties which are analogous to those in flat spacetime. For example the spacetime norm of the tangent vector is preserved. To check this compute

$$\frac{d}{d\lambda}(g_{\mu\nu}\dot{x}^\mu \dot{x}^\nu) = v^\rho \partial_\rho (g_{\mu\nu} v^\mu v^\nu) = v^\rho \nabla_\rho (g_{\mu\nu} v^\mu v^\nu) = 2 g_{\mu\nu}(v^\rho \nabla_\rho v^\mu) v^\nu = 0 \tag{5.30}$$

where we have used axiom (i) in the second equality sign, (ii,iv,v) in the third equality sign and (5.29) in the last. It follows in particular that a geodesic can not change its causal nature. Suppose a geodesic is timelike, i.e. $g_{\mu\nu} v^\mu v^\nu < 0$. We perform a reparametrization of the geodesic $\lambda \mapsto \bar{\lambda}$ in such a way that

$$\bar{v}^\mu = \frac{dx^\mu}{d\bar{\lambda}} = \frac{dx^\mu}{d\lambda} \frac{d\lambda}{d\bar{\lambda}} \tag{5.31}$$

satisfies $g_{\mu\nu} \bar{v}^\mu \bar{v}^\nu = -1$. It is now easy to see that this new parameter $\bar{\lambda}$ is nothing but proper time along the geodesic, or $d\bar{\lambda} = \sqrt{-g_{\mu\nu} v^\mu v^\nu} d\lambda = ds$.

Exercise 48: Consider the Lagrangian $L(z, \dot{z})$ given by $L = \frac{1}{2} g_{\mu\nu}(z) \dot{z}^\mu \dot{z}^\nu$. Calculate the corresponding Euler-Lagrange (EL-) expression. e.o.e.
Exercise 49: Let $z(\tau)$ be a curve with zero EL-expression. Show that L is constant along any such curve. Also try to infer this directly from the Lagrangian without knowing the EL-expression. e.o.e.

We now prove a result which is the mathematical analogue of the EP. Namely, let p be an event, given for simplicity by the coordinates $x^\mu = 0$. We can also assume the coordinates to be such that $g_{\mu\nu}(0) = \eta_{\mu\nu}$. We now show that there are new coordinates $\bar{x}^\mu$ near p, in which these two properties still hold, but in addition that $\bar{\Gamma}^{\mu}_{\nu\lambda}(0) = 0$. Such coordinates are called Riemannian normal coordinates (RNC's). The proof goes simply by checking that the transformation

$$x^{\mu} \mapsto \bar{x}^{\mu} = x^{\mu} + \frac{1}{2}\,\Gamma^{\mu}_{\nu\lambda}(0)\,x^{\nu}x^{\lambda} \tag{5.32}$$

does the job. Clearly p is still given by $\bar{x}^{\mu} = 0$ and $\bar{g}_{\mu\nu}(0) = \eta_{\mu\nu}$, since $\frac{\partial \bar{x}^{\mu}}{\partial x^{\nu}}(0) = \delta^{\mu}{}_{\nu}$. The rest follows by (5.25).
Suppose we write the geodesic equation at the point p in the 'barred' coordinate system from above. It then takes the form

$$\ddot{x}^{\mu} = 0 \tag{5.33}$$

Consequently, if we identify the 'force' acting on a freely falling particle as coming from the Christoffel term in the geodesic equation, we have shown that this can to leading order be eliminated by going to the 'freely falling system' given by (5.32).
The 'geodesic hypothesis', namely the assumption that freely falling particles move along geodesics, can be described as follows: take the equation of a free particle, i.e. one moving along a straight line in Minkowski space and replace ordinary derivatives by covariant derivatives.
We go further by asking how other physical systems, whose description is known in the absence of gravity, behave when a gravitational field is present. The answer to this question (which has to be postulated of course: the final judgment is spoken by experiment) is again: First take the special relativistic equations of motion, i.e. in Minkowski space. These contain partial derivatives of the basic quantities (usually first or second order). The rule then is to replace these partial derivatives by covariant derivatives with respect to the spacetime metric which describes the gravitational field. We repeat this rule in a way which is conceptually a little more precise: In the absence of gravity physical systems are described by certain equations on Minkowski space. These equations contain covariant derivatives with respect to the Minkowski metric (in other words: partial derivatives in standard Minkowskian coordinates, i.e. coordinates in which the Minkowski metric has its standard form $g_{\mu\nu} = \eta_{\mu\nu} = \mathrm{diag}(-1, 1, 1, 1)$). If we 'turn on' a gravitational field, we replace the Minkowski metric by whatever spacetime metric describes this gravitational field and replace the covariant derivatives with respect to the Minkowski metric by covariant derivatives with respect to this spacetime metric. Take electromagnetism for example, say in vacuum ($j^{\mu} = 0$). Then the Maxwell equations in the presence of gravity are simply given by

$$\nabla^{\nu}F_{\mu\nu} = 0 \qquad\qquad \nabla_{[\mu}F_{\nu\lambda]} = 0 \tag{5.34}$$

Exercise 50: Let $F_{\mu\nu}$ be a Maxwell field on a curved spacetime and let $F_{\mu\nu} =$

$-2\nabla_{[\mu}A_{\nu]}$ in the (generalized) Lorentz gauge $\nabla_\mu A^\mu = 0$. Show that

$$\Box A_\mu - R_\mu{}^\nu A_\nu = 0 \tag{5.35}$$

Similarly, for a perfect fluid in the presence of gravity, the energy momentum tensor is still given by

$$T^{\mu\nu} = (\rho + p)u^\mu u^\nu + p\, g^{\mu\nu} \tag{5.36}$$

with u^μ satisfying $g_{\mu\nu}u^\mu u^\nu = -1$, and the equations of motion follow from $\nabla_\nu T^{\mu\nu} = 0$.
Furthermore, in the case of dust $p = 0$ we still find that the integral curves of u^μ are geodesics, i.e. $u^\nu \nabla_\nu u^\mu = 0$, in accordance with the geodesic principle.

5.4 The concept of curvature

Suppose $g_{\mu\nu}$ is flat. Then, in suitable coordinates, the covariant derivative coincides with the partial derivative. Consequently, in that case, there holds that the operator $\nabla_{\mu\nu}$ given by $\nabla_\mu\nabla_\nu - \nabla_\nu\nabla_\mu$ vanishes on all tensor fields. In the general case the operator $\nabla_{\mu\nu}$ is not zero, but has still the remarkable property of being 'algebraic' in the following sense, for example in the case of covectors:

$$\nabla_{\mu\nu}(\phi\, \omega_\rho) = \phi \nabla_{\mu\nu}\, \omega_\rho \tag{5.37}$$

for every scalar field ϕ.

Exercise 51: Prove equation (5.37). e.o.e.

The point of (5.37) is that no derivatives of ϕ appear on the r.h. side. In other words, the l.h. side of (5.37) defines a linear map from covector fields to third-rank covariant tensors. In other words, there is a tensor $R_{\mu\nu\rho}{}^\sigma$ such that

$$(\nabla_\mu\nabla_\nu - \nabla_\nu\nabla_\mu)\, \omega_\rho = R_{\mu\nu\rho}{}^\sigma \omega_\sigma \tag{5.38}$$

This is the Riemann tensor. Given the knowledge of the Riemann tensor, (5.38) is called the Ricci identity. In the case of a covariant tensor of 2nd rank, one finds that

$$(\nabla_\mu\nabla_\nu - \nabla_\nu\nabla_\mu)\, t_{\rho\sigma} = R_{\mu\nu\rho}{}^\tau t_{\tau\sigma} + R_{\mu\nu\sigma}{}^\tau t_{\rho\tau} \tag{5.39}$$

This obtained e.g. by noting that every such tensor can be gotten by a linear combination of tensor products of covectors.
The Riemann tensor satisfies a number of identities. These are:

1. $R_{\mu\nu\rho}{}^\sigma = R_{[\mu\nu]\rho}{}^\sigma$

2. $R_{[\mu\nu\rho]}{}^{\sigma} = 0$

3. $R_{\mu\nu\rho\sigma} = R_{\mu\nu[\rho\sigma]}$

(For the definition of the l.h. side in the 2nd identity recall 4.9.) Property 1 is obvious. To prove property 2 observe first that (proof?)

$$\nabla_{[\mu}\nabla_{\nu}(f\,\nabla_{\rho]}g) = 0 \tag{5.40}$$

and then that every covector ω_μ is a sum of terms of the form $f\nabla_\mu g$. For property 3 it suffices to write down the evident identity

$$(\nabla_\mu\nabla_\nu - \nabla_\nu\nabla_\mu)\, g_{\rho\sigma} = 0 \tag{5.41}$$

We now show that the above properties imply a fourth one, namely

4. $R_{\mu\nu\rho\sigma} = R_{\rho\sigma\mu\nu}$

For the proof we use

$$\begin{aligned}
2R_{\mu\nu\rho\sigma} &= R_{\mu\nu\rho\sigma} + R_{\nu\mu\sigma\rho} = && (5.42)\\
&= -R_{\rho\mu\nu\sigma} - R_{\nu\rho\mu\sigma} - R_{\sigma\nu\mu\rho} - R_{\mu\sigma\nu\rho} = && (5.43)\\
&= -R_{\mu\rho\sigma\nu} - R_{\rho\nu\sigma\mu} - R_{\nu\sigma\rho\mu} - R_{\sigma\mu\rho\nu} = && (5.44)\\
&= +R_{\rho\sigma\mu\nu} + R_{\sigma\rho\nu\mu} = 2R_{\rho\sigma\mu\nu}\ , && (5.45)
\end{aligned}$$

where in the 2nd line we have used property 2 and in line 3 we have used properties 1 and 3. In the 4th line we have used property 2 on the 1st and 4th term and respectively on the 2nd and 3rd term.
How many independent components does the curvature tensor have at each point? First of all, by the symmetries 1,3 and 4 the number of is (in 4 dimensions) equal to $1 + 2 + 3 + 4 + 5 + 6 = 21$, with '6' arising as $1 + 2 + 3$ for the number of independent components of an antisymmetric 2-tensor in 4 dimensions. For property 2 note that the expression for $R_{[\mu\nu\rho]\sigma} = \frac{1}{3}(R_{\mu\nu\rho\sigma} + R_{\rho\mu\nu\sigma} + R_{\nu\rho\mu\sigma})$ is by virtue of the other symmetries already antisymmetric in all 4 indices. I t thus just suffices to add $R_{[\mu\nu\rho\sigma]}$ as remaining independent symmetry, and this, in 4 dimensions, amounts to just 1 more condition. So, finally the curvature tensor has 20 independent components. By a similar reasoning we find that the curvature tensor has 6 independent components in 3 dimensions.
There is also a differential identity, called Bianchi identity, obeyed by the curvature tensor namely

$$\nabla_{[\mu}R_{\nu\rho]\sigma\lambda} = 0 \tag{5.46}$$

For the proof, write the expression $2\nabla_{[\mu}\nabla_{\nu}\nabla_{\rho]}\omega_\sigma$ in two different ways, namely first as

$$2\nabla_{[\mu}\nabla_{[\nu}\nabla_{\rho]]}\omega_\sigma = \nabla_{[\mu}(R_{\nu\rho]\sigma}{}^{\tau}\omega_\tau) \tag{5.47}$$

and second as

$$2\nabla_{[[\mu}\nabla_{\nu]}\nabla_{\rho]}\omega_\sigma = R_{[\mu\nu\rho]}{}^\tau\nabla_\tau\omega_\sigma + R_{[\mu\nu|\sigma|}{}^\tau\nabla_{\rho]}\omega_\tau \tag{5.48}$$

where the vertical lines inside the antisymmetrization symbol on the right mean that that the index between these lines does not participate in the antisymmetrization. Comparing (5.47) with (5.48) as well as using Property 2 and the arbitrariness of ω_μ yields the desired result (5.46).
An important definition is that of the Ricci tensor $R_{\mu\nu}$ given by

$$R_{\mu\nu} = R_{\mu\rho\nu}{}^\rho \tag{5.49}$$

The contraction $R_{\mu\nu}g^{\mu\nu} = R$ is called Ricci scalar or scalar curvature. Using property 4 it is not difficult to check that the Ricci tensor is symmetric. We also define

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}\,g_{\mu\nu}R\,. \tag{5.50}$$

The tensor $G_{\mu\nu}$ is called Einstein tensor. We leave it as an easy exercise to show that the Bianchi identity implies the identity

$$\nabla^\mu G_{\mu\nu} = 0 \tag{5.51}$$

5.5 Curvature in terms of the metric

We can rewrite (5.38) in terms of ω_μ and, its first and second partial derivatives and the Christoffel symbols and their partial derivatives. We know in advance that all terms with derivatives of ω_μ have to drop out in the final relation. Thus, with the understanding that all such terms are omitted, we can write

$$\nabla_\nu\omega_\rho \triangleq -\Gamma^\tau_{\nu\rho}\omega_\tau \tag{5.52}$$

$$\nabla_\mu\nabla_\nu\,\omega_\rho \triangleq -(\partial_\mu\Gamma^\tau_{\nu\rho})\omega_\tau + \Gamma^\lambda_{\mu\nu}\Gamma^\tau_{\lambda\rho}\omega_\tau + \Gamma^\lambda_{\mu\rho}\Gamma^\tau_{\nu\lambda}\omega_\tau \tag{5.53}$$

Consequently there holds

$$R_{\mu\nu\rho}{}^\tau = -2\,\partial_{[\mu}\Gamma^\tau_{\nu]\rho} + 2\,\Gamma^\lambda_{\rho[\mu}\Gamma^\tau_{\nu]\lambda} \tag{5.54}$$

In particular the Riemann tensor, viewed as a functional of the metric, depends algebraically on the metric and on its first and second derivatives. It is also not difficult to see that this dependence is quasilinear, i.e. $R_{\mu\nu\rho}{}^\tau$ depends linearly on the highest, i.e. second, derivatives of $g_{\mu\nu}$.
Let the metric be flat metric, i.e. $g_{\mu\nu} = \eta_{\mu\nu}$. Then the Riemann tensor is zero (why?). Suppose next that the actual metric is only close to the flat one. Then, clearly, the Riemann tensor will be close to zero. Let $g_{\mu\nu} = \eta_{\mu\nu} + \epsilon h_{\mu\nu}$. We want to evaluate the Riemann tensor up to terms of order ϵ^2. Using (5.54) we find that

$$R_{\mu\nu\rho}{}^\tau = -\epsilon\eta^{\tau\sigma}(\partial_\rho\partial_{[\mu}h_{\nu]\sigma} - \partial_\sigma\partial_{[\mu}h_{\nu]\rho}) + O(\epsilon^2) \tag{5.55}$$

and consequently

$$R_{\mu\rho} = -\frac{\epsilon}{2}\left(\partial_\mu\partial_\rho h - 2\partial_\tau\partial_{(\mu}h_{\rho)}{}^\tau + \Box h_{\mu\rho}\right) + O(\epsilon^2)\,, \tag{5.56}$$

where the index on $h_{\mu\nu}$ is raised with $\eta^{\mu\nu}$ and $h \doteqdot \eta^{\mu\nu}h_{\mu\nu}$.

5.6 The meaning of curvature

In section 5.2 we defined the notion of flat space basically as 'Minkowski space in arbitrary coordinates'. It would be nice to have a tensorial characterization of flat metrics. As mentioned in the previous section, Minkowski space has zero curvature, and this latter condition is clearly tensorial. We state without proof that the vanishing of curvature is also sufficient in order for a metric to flat. Thus this is a tensorial characterization of flat metrics.
The presence of curvature is the sign of a true gravitational field. As the equivalence principle shows, there is no invariant concept of gravitational force. But there is, as we now show, an invariant concept of tidal force, and here the curvature tensor plays the essential role. Suppose we have a one parameter family $x^\mu(\epsilon; s)$ of point particles moving in a given gravitational field. Define $\mathring{x}^\mu(s) \doteqdot x^\mu(0;s)$ and $\delta x^\mu(s) \doteqdot \frac{d}{d\epsilon}|_{\epsilon=0}\, x^\mu(\epsilon; s)$. For each ϵ the path x^μ satisfies

$$\ddot{x}^\mu + \Gamma^\mu_{\nu\rho}(x)\,\dot{x}^\nu\dot{x}^\rho = 0 \tag{5.57}$$

Differentiating (5.57) with respect to ϵ and setting ϵ equal to zero, we obtain a linear second-order ODE for $\delta x^\mu(s)$ (with s-dependent coefficients) - the geodesic equation linearized at the solution $\mathring{x}^\mu(s)$, also called equation of geodesic deviation or Jacobi equation. It is given by

$$(\mathring{v}^\nu\nabla_\nu)(\mathring{v}^\rho\nabla_\rho)\,\delta x^\mu - R^\mu{}_{\nu\rho\sigma}(\mathring{x}(s))\mathring{v}^\nu\mathring{v}^\rho\delta x^\sigma = 0 \tag{5.58}$$

where $\mathring{v}^\mu(s) \doteqdot \frac{d}{d\epsilon}|_{\epsilon=0}\,\dot{x}^\mu(\epsilon; s)$. Deriving (5.58) directly is quite heavy. One can make life easier for example by working in RNC's based at the point $\mathring{x}^\mu(s)$ for some s, so that all Christoffel symbols are absent - but their derivatives are not. (Note that derivatives of the Christoffel symbols are present both in the Riemann tensor and the first term in (5.58).) At the end one replaces ordinary derivatives by covariant derivatives to obtain (5.58).
What is the Newtonian analogue of the Jacobi equation? To find this, we have linearize the Newtonian free-fall equation, namely

$$\ddot{x}^i = -\delta^{ij}\partial_j\Phi \tag{5.59}$$

where $x^i \in \mathbb{R}^3$, a dot denotes derivative with respect to absolute time and Φ is the gravitational potential. Now the recipe is the same: consider a 1-parameter family of solutions $x^i(\epsilon;t)$ and define $\delta x^i(t) \doteqdot \frac{d}{d\epsilon}|_{\epsilon=0}\, x^i(\epsilon;t)$. Then (5.59) implies

$$(\ddot{\delta x})^i(t) = -\delta^{ij}(\partial_j\partial_k\Phi)(\mathring{x}(t))\,\delta x^k(t) \tag{5.60}$$

So the Hessian of Φ plays a similar role in Newtonian theory as the Riemann tensor in GR.

5.7 Newtonian limit of the geodesic equation

We again have a small parameter ϵ and assume that $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$ where $h_{\mu\nu} = O(\epsilon^2)$. We take Minkowskian coordinates (t, x^i) with respect to $\eta_{\mu\nu}$ and use t to parameterize the geodesic path x^μ. We also suppose that $x^i(t) = O(1)$. Finally we assume that a derivative with respect to t makes every quantity smaller by $O(\epsilon)$, so for example $v^i(t) \doteqdot \frac{dx^i}{dt} = O(\epsilon)$. These assumptions together serve to make precise the idea that we are considering a particle moving slowly in a weak gravitational field. It remains to show that, with these assumptions, the geodesic equation has a consistent limit, and that this limit, which has to be $O(\epsilon^2)$ since $\frac{dv^i}{dt} = O(\epsilon^2)$, coincides with the Newtonian law of motion under an appropriate identification.
First of all the proper time s is related to t by

$$s(t) = \int^t \left(-g_{\mu\nu}\frac{dx^\mu}{dt'}\frac{dx^\nu}{dt'}\right)^{\frac{1}{2}} dt' \tag{5.61}$$

Since $\frac{dx^\mu}{dt'} = (1, v^i) = (1, O(\epsilon))$ and $h_{\mu\nu} = O(\epsilon^2)$, we infer that

$$s = t(1 + O(\epsilon^2)) \tag{5.62}$$

so that

$$\dot{v}^i + \Gamma^i_{00} = 0 \tag{5.63}$$

up to terms of order at least ϵ^3. Since $g^{\mu\nu} = \eta^{\mu\nu} + O(\epsilon^2)$ we have that

$$\Gamma^i_{00} = \frac{1}{2}\eta^{ij}(2\,\partial_t h_{j0} - \partial_j h_{00}) = -\frac{1}{2}\,\delta^{ij}\partial_j h_{00} + O(\epsilon^3) \tag{5.64}$$

Thus, with the identification $h_{00} = -2\Phi$ or, with correct dimensions,

$$h_{00} = -\frac{2\Phi}{c^2}\,, \tag{5.65}$$

our claim has been proved.

5.8 The Einstein field equations

We so far have postulated laws of motion for systems in a given gravitational field, but no laws yet which govern gravity itself. Recall the Newtonian law is

$$\Delta\Phi = 4\pi G\rho \tag{5.66}$$

with ρ being the mass density of matter and G the Newton constant.
Since object describing gravity is a Lorentzian metric $g_{\mu\nu}$, the field equations clearly have to be equations having this metric as the basic unknown. In addition we require this equation to be tensorial in nature. The simplest tensor one can build from the metric is the Ricci tensor. In fact, it is a theorem that there are no tensors just built from the metric and first derivatives, and the only such tensor depending on up to second derivatives of the metric and which is quasilinear has, up to an overall multiplicative constant, to be of the form

$$S_{\mu\nu} = R_{\mu\nu} + aRg_{\mu\nu} + \Lambda g_{\mu\nu} \tag{5.67}$$

for some constants a, b. The word 'just' in the above statement refers to the assumption that there should be no dependence on any other structure such as some give ('absolute') tensor field. This e.g. rules out that the scalars a, Λ are some given functions on spacetime rather than constants. When it is in addition required that $S_{\mu\nu}$ be divergence-free, this fixes a to be given by $a = -\frac{1}{2}$. Thus a reasonable guess for the field equations of general relativity seems to consist in setting $S_{\mu\nu}$ to be equal to a suitable constant κ times the stress energy tensor of whatever matter happens to be present, i.e.

$$S_{\mu\nu} = \kappa T_{\mu\nu} \tag{5.68}$$

Since equations of motion of the matter should presumably be given by $\nabla_\nu T^{\mu\nu} = 0$, this seems to be consistent in general only when $a = -\frac{1}{2}$. (When $a \neq -\frac{1}{2}$, then the matter equations together with (5.68) would imply that the scalar curvature is constant.)
What about the Λ-term in (5.68)? Well, this term - which has dimension $(\text{length})^{-2}$ - is the famous cosmological term whose stipulated presence Einstein once called his greatest blunder. At present most cosmologists believe that data in fact do require a non-zero (namely positive) value for Λ, but one which is very small ($\sim 10^{-46}\text{km}^{-2}$) on the length scale say of our own galaxy ($\sim 10^{18}$ km). Leaving out Λ for the time being, our candidate field equations take the form

$$G_{\mu\nu} = \kappa T_{\mu\nu} \tag{5.69}$$

or equivalently

$$R_{\mu\nu} = \kappa(T_{\mu\nu} - \frac{1}{2}g_{\mu\nu}T) + \Lambda g_{\mu\nu} \tag{5.70}$$

where $T \doteqdot g^{\mu\nu}T_{\mu\nu}$. (Derive (5.70) from (5.69) by taking the trace of (5.69).)

Next we will show that equation (5.69) has the correct Newtonian limit if κ is suitably chosen. From (5.56,5.65) we have, in the Newtonian limit, that

$$R_{00} = -\frac{1}{2}\,\Delta\left(-2\,\frac{\Phi}{c^2}\right) \tag{5.71}$$

Next recall that the $(0,0)$ - component of of the stress-energy tensor has the meaning of energy density, the other components are small at least of the order of the velocity of particles making up the material described by $T_{\mu\nu}$. E.g. for a perfect composed of nonrelativistic particles the pressure p has $p \ll \rho c^2$. Consequently

$$\Delta\Phi = \frac{\kappa}{2}\,c^4\rho \tag{5.72}$$

So that, with the identification

$$\kappa = \frac{8\pi G}{c^4} \tag{5.73}$$

we have the correct Newtonian limit. We will henceforth use 'geometrical' units, in which c and G are both equal to 1.
We end this section with a few remarks on the general nature of the equations (5.70) Although the Einstein equations have the correct Newtonian limit, the richness of physical phenomena they describe extends far beyond that of Newtonian theory. We list some such phenomena here.
(1) Existence of wave like solutions. There are no wave-like solutions in Newtonian gravity: the equation (5.66) describes action-at-a-distance incompatible with the principles of Special Relativity. Any time dependence $\rho(t,\vec{x})$ is instantly transferred to the field $\Phi(t,\vec{x})$. The same statement would not be true for electrodynamics or scalar field theory. Related to this is the following fact: suppose we look at solutions of (5.66) with no source and such that the field corresponds to an isolated system in the sense that Φ approaches a constant value (which can without loss be set to zero) at infinity. It then follows that Φ has to be zero everywhere. To prove this (see e.g. Mathematical Methods of Physics II), one multiplies the homogenous (5.66) by Φ and integrates over $\mathbb{R}^3$. Assuming $\Phi = O(\frac{1}{r})$ and $\vec{\nabla}\Phi = O(\frac{1}{r^2})$ at infinity and integrating by parts using the Gauss theorem, we end up with $\int_{\mathbb{R}^3}\left(\vec{\nabla}\Phi\right)^2 d^3x = 0$, which can only hold if Φ is constant whence zero.
What about gravity with no sources? There is clearly one solution of (5.69) with $T_{\mu\nu} = 0$, namely Minkowski spacetime, which has zero curvature whence zero Ricci tensor. One can next look at solutions which are small perturbations of Minkowski. According to (5.56) they are governed by

$$\Box h_{\mu\nu} + \partial_\mu\partial_\nu h - 2\partial_\tau\partial_{(\mu}h_{\nu)}{}^\tau = 0 \tag{5.74}$$

which - apart from the second and third term - has the character of a tensor wave equation on Minkowski space, i.e. when $\partial_\nu(h_\mu{}^\nu - \frac{1}{2}\delta_\mu{}^\nu h) = 0$, (5.74) takes the form $\Box h_{\mu\nu} = 0$. This difficulty apart, which is related to the coordinate invariance of the theory, we have a wave-type equation and correspondingly wave-type solutions. The direct detection of such gravitational waves is part of an international large-scale experimental effort under way at present.
(2) All forms of matter couple to gravity. Take source-free electromagnetism for example. Then the full (i.e. gravity plus electromagnetism) set of equations is given by the 'Einstein-Maxwell system', i.e.

$$G_{\mu\nu} = \frac{\kappa}{4\pi}\left[F_{\mu\rho}F_\nu{}^\rho - \frac{1}{4}\,\eta_{\mu\nu}\,F_{\rho\sigma}F^{\rho\sigma}\right] \tag{5.75a}$$

$$\nabla^\mu F_{\mu\nu} = 0 \tag{5.75b}$$

$$\nabla_{[\mu}F_{\nu\rho]} = 0 \tag{5.75c}$$

Since light rays follow null geodesics, they are affected by gravity (see light deflection in the Schwarzschild spacetime). It is thus entirely reasonable that the converse is also true, i.e. electromagnetism affects gravity. However due to the weakness of the gravitational interaction this latter effect is very weak under normal circumstances.
(3) The lack of a background structure. All theories of physics refer to some given geometric structure and some, like the old ether theory, require even a given frame of reference. Electromagnetism for example requires for its formulation the Minkowski metric of special relativity. This in turn does not single out a reference frame, bit it does single out a class of reference frames, namely the Lorentzian ones which (depending on the concept of reference frames which we have not made completely precise) are acted upon by the Lorentz or Poincaré group. No such structure is available in general relativity: the curved metric $g_{\mu\nu}$ is by no means given, but subject to the Einstein field equations. (And since these latter equations are tensorial, no frame of reference (coordinate system) is a priori singled out from all others). In a gravitational wave for example it is, strictly speaking, spacetime itself which is evolving rather than some field on a given background[1]. This unique feature of general relativity gives rise to a host of interesting conceptual issues. Sticking out among those problems is that of finding a quantum description of gravity ('quantization of the gravitational field').

[1]That statement notwithstanding there are many situations where the physical setup provides a background. E.g. in the previous paragraph we have mentioned small perturbations of Minkowski space, and these of course do refer to a background.

5.9 The Schwarzschild spacetime

The motion of a small body ('planet') in the gravitational field of a big body ('sun') is a basic problem of Newtonian physics. The first thing we need is the gravitational field itself for a spherical body at rest. In Newtonian physics, in the vacuum region, the general spherically symmetric solution of

$$\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{d\Phi}{dr}\right) = 0 \tag{5.76}$$

is of the form

$$\Phi = \frac{C}{r} + D \tag{5.77}$$

The constant D leaves the motion of the planet unaffected and is thus set to zero. The constant C is nothing but minus the mass $M = 4\pi \int \rho(r')r'^2 dr'$ of the sun. To prove this we consider

$$\int_{r=R} \vec{\nabla}\Phi \, d\vec{S} = -4\pi C = 4\pi \int \rho \, d^3x \tag{5.78}$$

where in the second equality we have used the Gauss theorem together with $\Delta\Phi = 4\pi\rho$ (recall we have set $G = 1$). The solution $\Phi = -\frac{M}{r}$ blows up at the origin. But that should not worry us: we expect the solution to be invalid below some nonzero radius R, which is where the surface of the central star is located.

The above calculation can be described by saying that there is a 1-parameter family of spherically symmetric solutions of $\Delta\Phi = 0$ on $\mathbb{R}^3\backslash\{\vec{0}\}$. The aim now is to study exactly the same problem for the vacuum Einstein equations. There arises an immediate question: what does it mean for a curved spacetime to be spherically symmetric? First recall the concept of 'symmetry': this means (see Sect.1.2 and Appendix B) a transformation $x \mapsto \bar{x} = f(x)$ not the identity leaving the metric invariant in the sense that

$$\frac{\partial f^\mu(x)}{\partial x^\rho}\frac{\partial f^\nu(x)}{\partial x^\sigma}\, g_{\mu\nu}(f(x)) = g_{\rho\sigma}(x) \tag{5.79}$$

Another interlude on Symmetries

Suppose we have a 1-parameter family of symmetries $f(\epsilon; x)$ near the identity, i.e. $f(0, x) = x$, satisfying Eq.(5.79). We define a vector field ξ^μ by

$$\xi^\mu(x) \doteq \frac{d}{d\epsilon}\, f^\mu(\epsilon; x)\Big|_{\epsilon=0} \tag{5.80}$$

Then (5.79) implies that

$$g_{\mu\nu,\rho}\xi^\rho + 2\,\xi^\rho{}_{,(\mu}g_{\nu)\rho} = 0 \tag{5.81}$$

The l.h. side of (5.81) is called the Lie derivative of the metric g with respect to the vector field ξ denoted by $(\mathcal{L}_\xi g)_{\mu\nu}$. There is the identity (proof?)

$$(\mathcal{L}_\xi g)_{\mu\nu} = \nabla_\mu \xi_\nu + \nabla_\nu \xi_\mu \tag{5.82}$$

The equation (5.81) is called the Killing equation. This consideration has a converse. Namely each solution of (5.81) gives rise to a 1-parameter family of symmetries (a 'continuous symmetry') as follows: consider the 'flow generated by ξ', i.e. the 1-parameter family of transformations $f(\epsilon; x)$ given by the solution to the initial value problem

$$\frac{d}{d\epsilon} f^\mu = \xi^\mu(f) \qquad f^\mu|_{\epsilon=0} = x^\mu \tag{5.83}$$

Provided that ξ satisfies (5.81), the solution $f^\mu(\epsilon; x)$ of (5.83) satisfies (5.79) for all values of ϵ.
For our next remark we require the concept of 'comoving coordinates'. Suppose we have a vector field which is nowhere zero. Then it is a fact (which we do not prove here) that there exists - at least in some neighborhood of each point - a coordinate system (x^0, x^i), in terms of which the vector field takes the form $v^\mu \partial_\mu = \partial_0$.
Suppose now we have a non-zero Killing vector for the metric $g_{\mu\nu}$. Then, in comoving coordinates, the metric is independent of the x^0 - coordinate. This easily follows from (5.81).
We end this interlude by considering the case where $g_{\mu\nu}$ is the Minkowski metric. Then, in Minkowskian coordinates, the Killing equations take the form

$$\partial_\mu \xi_\nu + \partial_\nu \xi_\mu = 0 \tag{5.84}$$

implying that

$$\partial_\rho \partial_\mu \xi_\nu + \partial_\rho \partial_\nu \xi_\mu = 0 \tag{5.85}$$

Thus

$$\partial_\rho \partial_\mu \xi_\nu + \partial_\rho \partial_\nu \xi_\mu + \partial_\nu \partial_\rho \xi_\mu + \partial_\nu \partial_\mu \xi_\rho - (\partial_\mu \partial_\nu \xi_\rho + \partial_\mu \partial_\rho \xi_\nu) = 2\partial_\rho \partial_\nu \xi_\mu = 0 \tag{5.86}$$

Thus $\partial_\mu \xi_\nu = N_{\mu\nu}$ is a constant antisymmetric tensor. Consequently

$$\xi_\mu = N_{\mu\nu} x^\nu + d_\mu \tag{5.87}$$

where d_μ is constant. When $N_{\mu\nu} = 0$, the corresponding flows are the translations $f^\mu(\epsilon; x) = x^\mu + \epsilon d^\mu$. When e.g. $N_{12} = -N_{21}$ are the only non-zero contributions to (5.87), the corresponding flow consists of spatial rotations around the x^3 - axis of the form $f^0(\epsilon; x) = x^0$, $f^1(\epsilon; x) = x^1 \cos N_{12}\epsilon + x^2 \sin N_{12}\epsilon$, $f^2(\epsilon; x) = -x^1 \sin N_{12}\epsilon + x^3 \cos N_{12}\epsilon$, $f^3(\epsilon; x) = x^3$. When $N_{01} = -N_{10}$ are the only non-zero elements, the flow consists of boosts in the x^1 - direction of the form $f^0(\epsilon; x) =$

$x^0 \cosh N_{01}\epsilon + x^1 \sinh N_{01}\epsilon,\ f^1(\epsilon;x) = x^0 \sinh N_{01}\epsilon + x^1 \cosh N_{01}\epsilon,\ f^2(\epsilon;x) = x^2,$ $f^3(\epsilon;x) = x^3$

We now return to the main theme of this section, which is that we are interested in solutions to the vacuum Einstein equations which have spherical symmetry. Compared to the Newtonian situation from above the problem is now that it is not a priori evident what 'spherical symmetry' is supposed to mean. Although a precise definition is beyond our scope here, we make some remarks:

The requirement of spherical symmetry means that there are three Killing vectors which in some precise sense mimic the three Killing vectors obtained from (5.87) by setting $d_\mu = 0$ and $N_{0i} = 0$. For example these Killing vectors should be tangential to a family of 2-surfaces with the geometry of 2-dimensional spheres (corresponding to the spheres in Minkowski space given by $t = \text{const}$, $r = \text{const}$). One also has to assume that the group of motions generated by the Killing vectors is the rotation group $SO(3)$. Making a long story short, one finds that there exist coordinates (t, r, Θ, ϕ), so that the metric takes the form

$$g_{\mu\nu} dx^\mu dx^\nu = -V dt^2 + W dr^2 + r^2 (d\Theta^2 + \sin^2\Theta\ d\phi^2)\,, \tag{5.88}$$

where V, W depend only on (t, r). Note we are here using the notation introduced in Sect.(1.3). Note also that the flat metric corresponds to $V = W = 1$. One now has to compute, for the metric given by (5.88), to compute the Ricci tensor and set it to zero. After some lengthy computations one finds that the vacuum Einstein equations with the ansatz (5.88) have a one parameter class of solutions which, with some slight adjustment of coordinates can be written as

$$g_{\mu\nu} dx^\mu dx^\nu = -(1 - \frac{2C}{r})\, dt^2 + \frac{1}{1 - \frac{2C}{r}}\, dr^2 + r^2 (d\Theta^2 + \sin^2\Theta\ d\phi^2)\,, \tag{5.89}$$

for some constant C. This is the famous Schwarzschild family of metrics. Note that this metric is independent of the coordinate t and consequently the vector field ξ given by $\xi = \partial_t$ is a Killing vector. Thus with spherical symmetry the Einstein equations give us an additional symmetry 'for free'. Since the Schwarzschild metric is the metric outside a spherically symmetric body, this means that a pulsating spherical source will still create a gravitational field in its exterior which is given by (5.89), in particular which is spherically symmetric.

We have yet to determine the nature of the constant C. Writing $C = \frac{GM}{c^2}$ and performing the Newtonian limit we find from (5.65) that M plays exactly the role of the mass of the central body.

As written, the Schwarzschild metric makes sense for large radii, but breaks down at $r = R_s = \frac{2GM}{c^2}$. For many astrophysical bodies this is no problem: the Schwarzschild metric is only valid in the vacuum region, but not the matter region, and the Schwarzschild radius R_s can be well within the matter region. For the earth, for example, we have that $R_s = 1\text{cm}$. But, as we now know, the Schwarzschild metric can be given a meaning ('be extended') even at and inside the Schwarzschild radius, namely as a black hole. It is believed that there are plentiful objects in the universe which are realizations of the extended Schwarzschild metric.

5.10 Gravitational Redshift

Suppose we have two observers in the SS spacetime, both of which are at rest w.r. to the Killing vector $\xi = \partial_t$, e.g. at spatial locations (r_1, Θ_1, ϕ_1) and (r_2, Θ_2, ϕ_2) respectively. Their world lines, when parametrized by proper time, are thus integral curves of the vector field

$$u^\mu = \frac{\xi^\mu}{\sqrt{-(\xi, \xi)}} \tag{5.90}$$

Now consider two observers following integral curves of ξ at different values of (ξ, ξ), i.e. different r (the values of the angular coordinates do not matter), connected by light rays, i.e. curves with tangent k^μ where $(k, k) = 0$ and $k^\nu \nabla_\nu k^\mu = 0$. By a computation done in the following section the quantity (k, ξ) is conserved along light rays. Thus, in obvious notation, $(k, \xi)_1 = (k, \xi)_2$, so that

$$\frac{(u, k)_2}{(u, k)_1} = \sqrt{\frac{(\xi, \xi)_1}{(\xi, \xi)_2}} \tag{5.91}$$

But the left hand side of (5.91) is nothing but the ratio $\frac{\omega_2}{\omega_1}$ of the frequencies measured by the two observers.

5.11 Geodesics of Schwarzschild

We now have the task of studying the orbits of test particles in the gravitational field given by the Schwarzschild metric. Before doing so we explain the relationship between symmetries and conserved quantities, in the context of the geodesic equation. Suppose we have a solution to the geodesic equation, i.e. a path $x^\mu = x^\mu(\lambda)$, whose tangent vector $v^\mu = \frac{dx^\mu}{d\lambda}$ satisfies

$$v^\nu \nabla_\nu v^\mu = 0 \tag{5.92}$$

Suppose next that the spacetime metric entering Eq.(5.92) (via the Christoffel symbols) has a Killing vector field k^μ. Then, we claim, the quantity $Q(k)$ given by $Q = v^\mu k_\mu$ is conserved along the geodesic orbit. To prove this we compute

$$v^\mu \nabla_\mu Q = k_\mu (v^\nu \nabla_\nu v^\mu) + v^\mu (v^\nu \nabla_\nu k_\mu) = k_\mu \cdot 0 + v^\mu v^\nu \nabla_{(\nu} k_{\mu)} = 0 \tag{5.93}$$

Here have used the geodesic equation in the first equality sign. In the second equality sign we have used that the contraction between a symmetric tensor (in this case $a^{\mu\nu} = v^\mu v^\nu$) and an antisymmetric tensor (in this case $b_{\mu\nu} = \nabla_{[\nu} k_{\mu]}$) is zero, since $a^{\mu\nu} b_{\mu\nu} = -a^{\nu\mu} b_{\nu\mu} = -a^{\mu\nu} b_{\mu\nu} = 0$. In the last equality sign in (5.93) we have of course used the Killing equation.

Now back to the Schwarzschild metric. It is intuitively clear that, by spherical symmetry, we can without loss of generality assume that the spatial part of a geodesic orbit lies in the plane $\Theta = \frac{\pi}{2}$. Here is a somewhat more detailed argument: Recall that initial data for a geodesic orbit consist of an 8-tuple $(t_0, r_0, \Theta_0, \phi_0; \dot{t}_0, \dot{r}_0, \dot{\Theta}_0, \dot{\phi}_0)$, subject to the constraint

$$-\left(1 - \frac{2M}{r_0}\right) \dot{t}_0^2 + \frac{\dot{r}_0^2}{1 - \frac{2M}{r_0}} + r_0^2 \left(\dot{\Theta}_0^2 + \sin^2\Theta_0 \, \dot{\phi}_0^2\right) = \kappa \,, \tag{5.94}$$

where $\kappa = -1$, 0 depending on whether we are studying timelike or null geodesics. We can, by a suitable rotation, arrange that $\Theta_0 = \frac{\pi}{2}$ and $\dot{\Theta}_0 = 0$. Since the Schwarzschild metric is invariant under the 'equatorial map' $\Theta \mapsto \pi - \Theta$, the geodesic evolving from the initial data has to 'remain on the equator'. Thus $\Theta(s) \equiv \frac{\pi}{2}$. There is consequently no loss of generality to assume, as we now will, our geodesics to move on the equator. We next exploit the symmetry of the spacetime under $\xi = \frac{\partial}{\partial t}$ and $\eta = \frac{\partial}{\partial \phi}$. From what we learnt at the beginning of this section it follows that $-E = g_{\mu\nu}(x(\lambda))\dot{x}^\mu(\lambda)\xi^\nu(x(\lambda))$ and $l = g_{\mu\nu}(x(\lambda))\dot{x}^\mu(\lambda)\eta^\nu(x(\lambda))$ are both conserved along geodesics. We first confine attention to the timelike case. Then note that E is nothing but the particle's conserved energy (or rather energy per mass). In nonrelativistic terms this energy consists the rest energy, kinetic energy and potential energy. More explicitly E is given by

$$E = \left(1 - \frac{2M}{r}\right) \dot{t} \tag{5.95}$$

or, solving (5.94) with $\Theta = \frac{\pi}{2}$ and assuming $\dot{t}$ to be positive (i.e. the particle 'moves into the future' rather than past)

$$E = \sqrt{1 - \frac{2M}{r} + \dot{r}^2 + r^2 \left(1 - \frac{2M}{r}\right) \dot{\phi}^2} \tag{5.96}$$

The quantity l is given by

$$l = r^2 \dot{\phi} \tag{5.97}$$

which has the same form as the z component of the nonrelativistic angular momentum. Inserting (5.97) into (5.96), there results

$$E = \sqrt{1 - \frac{2M}{r} + \dot{r}^2 + \left(1 - \frac{2M}{r}\right) \frac{l^2}{r^2}} \tag{5.98}$$

When $\dot{r}^2$, $\dot{\phi}^2$ and M are all small, (5.98) yields

$$E \sim 1 + \frac{1}{2}\left(\dot{r}^2 + \frac{l^2}{r^2}\right) - \frac{M}{r} \tag{5.99}$$

which rest-mass energy plus Newtonian kinetic energy plus Newtonian potential energy, as promised.
Actually we will use the squared form of (5.99), written as

$$\frac{E^2 - 1}{2} = \frac{\dot{r}^2}{2} + V_l(r) \tag{5.100}$$

where

$$V_l(r) = -\frac{M}{r} + \left(1 - \frac{2M}{r}\right)\frac{l^2}{2r^2} \tag{5.101}$$

It follows from (5.100) that $\dot{r}$ can be expressed in terms r, so that any solution to the geodesic equation can be written in terms of quadratures. Thus the motion of test particles in the Schwarzschild spacetime is 'completely integrable'. In fact, like in any such system, it often unnecessary to even perform this integration by analytic or numerical means, since many qualitative properties can be read off from the form of the effective potential $V_l(r)$. To this we turn after the following exercise.

Exercise 52: Consider the radial ($l = 0$) timelike geodesic $r(s)$ in Schwarzschild which starts at $r(0) = R$ with $\dot{r}(0) = 0$. Verify that $(r(z), s(z))$ with

$$r = R\cos^2\frac{z}{2} \qquad\qquad s = \frac{1}{2}\left(\frac{R^3}{2M}\right)^{\frac{1}{2}}(z + \sin z)\ , \tag{5.102}$$

where $z \in [0, \pi)$, gives a parameter representation of the solution. What happens at $r = 2M$? How long does it take the particle to reach $r = 0$? e.o.e.

First recall for comparison the Newtonian effective potential, given by

$$V_l^{\text{newt}}(r) = -\frac{M}{r} + \frac{l^2}{2r^2} \tag{5.103}$$

It has the property that V_l goes to zero as r goes to ∞. For $l \neq 0$, i.e. nonradial motion $\dot{\phi} \neq 0$, it goes to infinity as r approaches zero and has a strict (negative) minimum at $r = \frac{l^2}{2M}$: this corresponds to circular orbits. All other orbits with $V_l(r) < 0$ oscillate between a minimal and maximal radius. The corresponding functions $r(t)$ can only be evaluated numerically, but their shapes $r(\phi)$ can be found explicitly (in fact: this was already done by Newton) and turn out to be the ellipses found observationally by Kepler. In particular these orbits are periodic. Orbits with $V_l(r) > 0$ are scattering orbits of hyperbolic shape. When $l = 0$, we get purely radial motions which for t going to either $-\infty$ or $+\infty$ or both approach the singularity at $r = 0$. Of course this should not worry us: the singularity is 'unphysical' anyway, it will be inside the matter region. The early relativists including Einstein thought similarly about the singularity of the

Schwarzschild metric at $r = R_s$. As it turns out, they were wrong.
The most conspicuous difference between V_l and $V_l^{\rm newt}$ is that there is no potential barrier in the former for $r \leq 2M$, so that V_l goes to $-\infty$ even when $l \neq 0$. Moreover, when $l^2 < 12M^2$, the effective potential is monotonically increasing in r. Thus no stable bound orbits exist for this range of parameters.
When $l^2 > 12M^2$ (this e.g. applies to the earth-sun system), there is a maximum at $r = r_-$ and a minimum at $r = r_+$, where $r_\pm$ satisfy $r^2 - r\frac{l^2}{M} + 3l^2 = 0$. There is thus in this case a range of values for $E^2 - 1$ such that the corresponding solutions have two branches, one of which has solutions $r(\lambda)$ oscillating between a minimal radius $r_{\rm min} > r_-$ and a maximal radius $r_{\rm max} < r_+$ (hint: picture the graph of $V_l(r)!)^2$. Of course there is the limiting case of solutions with $\frac{1}{2}(E^2 - 1) = V(r_+)$: these are stable circular orbits.

Exercise 53*: Consider a particle in the Schwarzschild geometry on a geodesic trajectory which is spatially spherical, i.e. $r = R = \text{const} > 3M$. Compare s, the proper time per revolution of that particle, with the elapsed time τ of a static observer at $r = R$ to show that

$$\frac{\tau}{s} = \sqrt{\frac{1 - \frac{2M}{R}}{1 - \frac{3M}{R}}} \tag{5.104}$$

The above oscillatory solutions are the analogue of the Kepler ellipses in Newtonian theory. However the function $r(\phi)$, although periodic, takes an angle greater than 2π to reach the same value of r: this is the famous phenomenon of perihelion advance. Except in the very special case, where the perihelion advances by a multiple of 2π (no perihelion advances as large as this have so far been observed), these orbits are not periodic in space. Of course, even in the Newtonian theory, orbits of say planets in the solar system are not exactly elliptical: the sun's gravitational field is not exactly spherically symmetric. More importantly there are perturbations on the orbit of each planet due to the presence of the other planets. In the case of Mercury, the innermost planet in our solar system, when these effects were taken into account, there was in the Newtonian prediction a residual advance of the perihelion amounting to 43 arcseconds per century, which remained unaccounted for. When Einstein realized, that general relativity in fact explains this effect, this was, according to him and very understandably, 'the happiest day of his life'. By now we know astrophysical systems with a much higher perihelion advance of relativistic origin. The binary pulsar PSR 1913+16 has a perihelion advance of 4 degrees per year, the same amount in a day as Mercury in a century.

[2] The other branch corresponds to solutions all of which enter the region $r < R_s$. There are no stable bound orbits.

We now outline a method to obtain the exact form of the perihelion advance for orbits close to a spherical one. We cast (5.100) in the form

$$\frac{\dot{r}^2}{2} + W_l(r) = \epsilon^2 \tag{5.105}$$

where $W_l(r) = V_l(r) - V_l(r_+)$, where r_+ is the larger solution of the equation $V_l'(r) = 0$. Clearly, when $\epsilon = 0$, this gives a stable circular orbit at $r = r_+$. We can write $\dot{r} = \frac{dr}{d\phi}\frac{l}{r^2}$. Inserting this into (5.105) we find that for the angle $\Delta\phi$ swept out between the turning points r_1 and r_2 to be

$$\Delta\phi = \frac{l}{\sqrt{2}} \int_{r_1}^{r_2} \frac{d\rho}{\rho^2\sqrt{\epsilon^2 - W_l(\rho)}} \tag{5.106}$$

Here $r_1 < r_+ < r_2$, where $W_l(r_{1,2}) = \epsilon^2$. Since $W_l(r) = \frac{(r-r_+)^2}{2} W_l''(r_+) + O((r - r_+)^3)$, $r_{1,2}$ are approximately given by ($\epsilon > 0$)

$$r_{1,2} \sim r_+ \mp \epsilon\sqrt{\frac{2}{W_l''(r_+)}} \tag{5.107}$$

and

$$W_l(r) - \epsilon^2 \sim \frac{V_l''(r_+)}{2}(r - r_1)(r - r_2) \tag{5.108}$$

Thus, in the limit that ϵ goes to zero, we obtain

$$\Delta\phi = \frac{l}{r_+^2\sqrt{V_l''(r_+)}} \int_{r_1}^{r_2} \frac{d\rho}{\sqrt{(r_2-\rho)(\rho-r_1)}} = \frac{l\,\pi}{r_+^2\sqrt{V_l''(r_+)}} \tag{5.109}$$

Using the explicit form of V_l given in (5.101) we easily find

$$\Delta\phi = \pi\left(\frac{1+\sqrt{1-\frac{12}{x^2}}}{1-\frac{12}{x^2}+\sqrt{1-\frac{12}{x^2}}}\right)^{\frac{1}{2}} \tag{5.110}$$

where $x = \frac{l}{M}$. Thus, to lowest order in $\frac{GM}{lc}$, we have

$$\Delta\phi \sim \pi + \frac{3\pi M^2 G^2}{l^2 c^2} \tag{5.111}$$

This ends our discussion of timelike geodesics in the Schwarzschild spacetime. We now turn to null geodesics. It is another important prediction of General Relativity that in the field of a spherically symmetric mass a light ray coming from infinity and again escaping to infinity gets deflected towards the central

mass.
Now the equation for $r(\phi)$ takes the form

$$\left(\frac{dr}{d\phi}\right)^2 \frac{l^2}{r^4} + \left(1 - \frac{2M}{r}\right)\frac{l^2}{r^2} = E^2 \tag{5.112}$$

Replacing the dependent variable r by $v = \frac{M}{r}$ and defining $\beta = \frac{ME}{l}$, the equation (5.112) acquires the form

$$\left(\frac{dv}{d\phi}\right)^2 + v^2 - 2v^3 = \beta^2 \tag{5.113}$$

In the absence of gravity the v^3 - term on the left in (5.113) would be missing and the solution would be

$$\frac{1}{r(\phi)} = \frac{E}{l}\sin(\phi - \phi_0) \tag{5.114}$$

which of course is a straight line in spherical coordinates. The angles ϕ_0 and $\phi_0 + \pi$ are the asymptotic directions of the light ray and $\frac{l}{E}$ plays the role of the impact parameter. Thus the constant β^2 in (5.113) is $[(\text{impact parameter})/M]^{-2}$.
The solutions can again be discussed in terms of the 'potential' f given by

$$f(v) = v^2 - 2v^3 \quad \text{for } v > 0 \tag{5.115}$$

The function f is monotonically increasing between the value 0 at $v = 0$ and its maximum value $\frac{1}{27}$ at $v = \frac{1}{3}$, after which it decreases monotonically. Thus for each value $0 \leq \beta^2 < \frac{1}{27}$ there is exactly one number v_0 in the interval $[0, \frac{1}{3})$ so that $f(v_0) = \beta^2$. The corresponding solutions $r(\phi)$, like the one given by (5.114), describe light rays coming in from infinity ($v = 0$), reaching a minimum distance from the central mass at radius $r_0 = \frac{M}{v_0}$ and escaping to infinity again. The angle Φ swept out during this process is given by

$$\Phi(v_0) = 2\int_0^{v_0} \frac{dv}{\frac{dv}{d\phi}} = 2\int_0^1 \frac{dt}{\sqrt{1 - t^2 - 2v_0(1 - t^3)}}\,, \quad 0 \leq v_0 < \frac{1}{3} \tag{5.116}$$

The quantity Φ is an analytic, monotonically increasing function in $[0, \frac{1}{3})$ with minimum value $\Phi(0) = \pi$, as it has to be. Since $\lim_{v \to v_0} f(v) = \infty$, the number of times which a light ray revolves around the central mass can be arbitrarily large. We can explicitly evaluate Φ for small v_0 by Taylor expanding around $v_0 = 0$. It turns out that

$$\left.\frac{d\Phi}{dv_0}\right|_{v_0=0} = 4\,, \tag{5.117}$$

so that

$$\Phi = \pi + \frac{4\,GM}{r_0\,c^2} + O\left(\frac{G^2M^2}{r_0^2\,c^2}\right) \tag{5.118}$$

Equation (5.118) was first experimentally verified for sunlight grazing the earth on a famous expedition to South America in 1919 led by the English astrophysicist Eddington. The news of this event instantly made Einstein a world celebrity.

5.12 Relativistic cosmology

5.12.1 Cosmological Principle

Cosmology is the study of the universe as a whole. Since space and time according to GR are by their very nature dynamical objects, GR has interesting predictions to make concerning the evolution of the universe. These however rest on some assumptions concerning the matter content of the universe, for which there is some incomplete empirical evidence. The main assumption is the so-called 'cosmological principle' which states that the universe be 'spatially homogenous and isotropic'. Namely by homogeneity our place in the universe should not be singled out from any other location: this assumption, which taken literally is of course blatantly wrong, seems in fact to be true 'on average', if this average is taken over distances of the order of several 100 Mpc. Isotropy means that no spatial direction, in the rest frame of the averaged-out matter (the 'cosmological substratum') should should be singled out over any other: the cosmic microwave background, an ideally thermal radiation of 2,7 degrees Kelvin first detected by Penzias and Wilson in 1964 is in fact isotropic to within 1 part in 10^4. Homogeneity then implies that the universe is also isotropic at any spatial point other than 'ours'.

How are we going to describe the geometry of such a universe? We have the four velocity field u^μ of the cosmic substratum subject to $g_{\mu\nu}u^\mu u^\nu = -1$. We claim isotropy implies that $\nabla_{[\mu}u_{\nu]} = 0$. Otherwise there would exist an antisymmetric, i.e. 'Maxwell-type', tensor field $F_{\mu\nu}$, which we could, relative to the rest frame defined by u^μ, decompose $F_{\mu\nu}$ into its electric and magnetic part, and if any of those would be zero, this would pick out a direction thus violating isotropy. Next observe that

Exercise 54: The tensor $\nabla_\mu u_\nu = \nabla_{(\mu}u_{\nu)}$ is 'purely spatial' in the sense that $u^\mu \nabla_\mu u_\nu = 0$. e.o.e.

At the same time Exercise 54 says that the vector field u^μ is geodesic.

The spatial symmetric tensor $a_{\mu\nu} = \nabla_\mu u_\nu$ has real eigenvalues λ determined by the equation $a_{\mu\nu}l^\nu = \lambda h_{\mu\nu}l^\nu$, where $h_{\mu\nu}$ is the spatial metric defined by $h_{\mu\nu} = g_{\mu\nu} + u_\mu u_\nu$. (Compare with $\Pi_{\mu\nu}$, where $\Pi^\mu{}_\nu$ is defined in (2.5).) But if any two of these three real numbers would be different, that would again give rise to an anisotropy. We thus conclude that there is a scalar field H, such that

$$\nabla_\mu u_\nu = H(g_{\mu\nu} + u_\mu u_\nu) \tag{5.119}$$

Again by isotropy, H can only depend on time, not space. Thus

$$\nabla_\mu H = -u_\mu \dot{H}\,, \tag{5.120}$$

where $\dot{H} = u^\mu \nabla_\mu H$. As we will see, H is nothing but the 'Hubble-constant', which in fact is not constant in time.

Given a non-zero vector field v^μ, there is always a choice of coordinates which are 'comoving' with this vector field, in other words coordinates (t, x) such that $u^\mu \partial_\mu = \partial_t$. This intuitively plausible fact is proved in texts on differential geometry. Assume this having been done with the vector filed u^μ. As was seen in Exercise 47, these coordinates are not unique, but for example can be changed by setting $\bar{x}^i = x^i$ and $\bar{t} = t - F(x^j)$. Next observe, that, by virtue of $\nabla_\mu u_\nu = \nabla_\nu u_\mu$, there exists a scalar field τ, such that $u_\mu = -\nabla_\mu \tau$. Now contracting with u^μ, we see that $u^\mu \nabla_\mu \tau = \dot{\tau} = 1$. Consequently the function τ is exactly of the form $\tau = t - F(x^i)$. Thus we can choose τ as our new time coordinate and, calling τ again t, still retain $u^\mu \partial_\mu = \partial_t$. In this coordinate system we also have that $u_\mu dx^\mu = -dt$, and since $u_\mu = g_{\mu\nu} u^\nu$, the metric has to satisfy that $g_{00} = -1$ and $g_{i0} = 0$. Consequently the 4-dimensional line element ds^2 is now of the form

$$ds^2 = -dt^2 + g_{ij}(t, x^k) dx^i dx^j \tag{5.121}$$

In the present coordinates we have that $\nabla_i u_j = \partial_i u_j - \Gamma^\mu_{ij} u_\mu = \Gamma^0_{ij} = \frac{1}{2} \dot{g}_{ij}$. Thus, using (5.119), there results

$$\dot{g}_{ij} = 2H g_{ij} \tag{5.122}$$

Spatial volumes V are given by the quantity $\int_B g^{\frac{1}{2}} d^3x$ over some comoving domain B in x- space, where $g = \det(g_{ij})$. Using the identity $\dot{g} = g g^{ij} \dot{g}_{ij}$, we find

$$\frac{\dot{V}}{V} = 3H\,, \tag{5.123}$$

or H measures relative changes of spatial length. So for an expanding universe, like the one we are in, the Hubble constant is positive.

Before continuing it will be convenient to rewrite H in terms of a characteristic length scale defined by

$$H = \frac{\dot{R}}{R} \tag{5.124}$$

This can always be done by defining $R(t) = \exp \int_{t_0}^t H(t')dt'$. Note $R(t)$ is unique up to a multiplicative constant c. It then follows from (5.122) that

$$(R^{-2} g_{ij})^{\dot{}} = 0 \tag{5.125}$$

Next note that the Ricci tensor of any metric is unchanged, when this metric g is rescaled by a constant factor, i.e. replaced by $c^2 g$, where c is a constant. Thus the Ricci tensor $\mathcal{R}_{ij}$ of the spatial metric g_{ij} remains unchanged, when we replace g_{ij} by a time-dependent factor, and the Ricci scalar $\mathcal{R}$ of g_{ij} changes by the power -2 of this factor. It thus follows from (5.125) that $R^2\mathcal{R}$ does not change with time. From the cosmological principle, and also from the computation we will do in Sect.(5.11.3), it follows that $\mathcal{R}$ is constant in space. We can thus, when $\mathcal{R}$ is non-zero, in an intrinsic way get rid of the scale freedom in the choice of $R(t)$ by requiring that

$$\mathcal{R} = \frac{6\,k}{R^2} \tag{5.126}$$

where $k = \pm 1$ dependent on the sign of $\mathcal{R}$, and continue using (5.126) also when $\mathcal{R}$ vanishes, by allowing also $k = 0$. In the latter case of course the scale freedom in $R(t)$ remains.

5.12.2 The cosmological red-shift

Suppose have two world lines of the cosmological substratum, say galaxies 1 and 2, and a light ray with tangent k^μ intersecting both. We want to compute $\omega' = k^\mu \nabla_\mu \omega$, where $\omega = -(k,u) = -g_{\mu\nu}k^\mu u^\nu$ and $' = \frac{d}{d\lambda}$ is derivative w.r. to the affine parameter associated with k^μ. We find

$$-\omega' = (k^\mu \nabla_\mu k^\nu)u_\nu + k^\mu k^\nu \nabla_\mu u_\nu = 0 + H k^\mu k^\nu h_{\mu\nu} = H\omega^2\,, \tag{5.127}$$

where we have used the geodesic equation for k^μ and Eq.(5.119) in the second equality and in the third equality have used the definition of $h_{\mu\nu}$ and the fact that k^μ is a null vector. Next we want to rewrite the (affine) '-derivative in terms of the $\dot{}$ - derivative $\frac{d}{dt}$. To do this we observe that $k^0 = \frac{dt}{d\lambda}$, and since $u^\mu = (1, \vec{0})$, it follows that $\frac{d}{dt} = \frac{1}{\omega}\frac{d}{d\lambda}$, so that

$$\dot{\omega} + H\omega = 0\,, \tag{5.128}$$

or, using (5.124), that

$$\dot{\omega}R + \omega\dot{R} = 0 \tag{5.129}$$

Eq.(5.129) in turn is the same as $(\omega R)\dot{} = 0$ or

$$\frac{\omega(t_2)}{\omega(t_1)} = \frac{R(t_1)}{R(t_2)} \tag{5.130}$$

In particular, when $H > 0$ ('galaxies recede from each other') we have redshift.
The equation (5.130), despite its fundamental nature, has to be interpreted with some care for two reasons. First one should realize that the two times t_1 and t_2 refer to emission, resp. absorption of a light ray on two different orbits of u^μ, i.e. two different points-in-space. In other words, t_1 is the retarded time along galaxy 1 for time t_2 along galaxy 2. This issue is taken care of by solving, given the knowledge of $R(t)$, the equations for null geodesics connecting the two world lines in question, that-is-to-say given a spatial distance l between them. To leading order in l one of course finds that

$$z = \frac{\omega_1}{\omega_2} - 1 = Hl\,, \tag{5.131}$$

which is the famous redshift-distance relation. The second, more serious, problem consists in noticing that the distance l is not a physical observable and should be replaced by some. For this there have been several proposals, but the actual measurement of distances over a range of thousands of Mpc's has and continues to be a great challenge for observational cosmology.

5.12.3 The dynamics of the universe

The evolution of the universe which in our case is the same as the time dependence of g_{ij}, is governed by the Einstein equations, which we write in the form

$$R_{\mu\nu} = 8\pi G(T_{\mu\nu} - \frac{1}{2}\, g_{\mu\nu}T) + \Lambda g_{\mu\nu} \tag{5.132}$$

We take the energy momentum tensor to be that of a perfect fluid and keep the Λ-term, since most cosmologists at present believe that the observational data can not be explained without a positive cosmological constant. Then (5.132) takes the form

$$R_{\mu\nu} = 8\pi G[(\rho + p)u_\mu u_\nu + \frac{1}{2}(\rho - p)g_{\mu\nu}] + \Lambda g_{\mu\nu} \tag{5.133}$$

or also

$$G_{\mu\nu} = 8\pi G[(\rho + p)u_\mu u_\nu + p\, g_{\mu\nu}] - \Lambda g_{\mu\nu} \tag{5.134}$$

For computing the $(0i)$ component of (5.132) we use the identity

$$\nabla^\nu \nabla_\mu u_\nu = R_\mu{}^\nu u_\nu + \nabla_\mu \nabla^\nu u_\nu \tag{5.135}$$

following from the Ricci identity. We now use (5.133) in the first term on the right of (5.135) and (5.119) for the remaining terms. We find that

$$3H^2 u_\mu = -4\pi G(\rho + 3p)u_\mu + \Lambda u_\mu - 3\dot{H} u_\mu \tag{5.136}$$

or

$$\dot{H} + H^2 = -\frac{4\pi G}{3}(\rho + 3p) + \frac{\Lambda}{3} \tag{5.137}$$

This is the famous Friedmann equation. It remains to consider the (ij) - components of the Einstein equations. For this purpose we need an identity which relates the spatial components of the curvature of the metric $g_{\mu\nu}$ with the curvature of g_{ij} for fixed t. This identity can be found by a calculation, which we skip. The answer is

$$R_{ijkl} = \mathcal{R}_{ijkl} + 2H^2 g_{k[i} g_{j]l} \tag{5.138}$$

We now trace (5.138) with $g^{ik} g^{jl}$. For the left-hand side we use the following algebraic identity

$$R_{ijkl} g^{ik} g^{jl} = 2G_{00} \tag{5.139}$$

Exercise 55: Prove identity (5.139). e.o.e.
We then find, using (5.134), that

$$2(8\pi G\rho + \Lambda) = \mathcal{R} + 6H^2 \tag{5.140}$$

But then, by virtue of (5.126), we have that $R^2(3H^2 - 8\pi G\rho - \Lambda)$ has zero time derivative. Calculating this time derivative and using the Friedmann equation we obtain

$$(\rho R^3)^{\cdot} + p\,(R^3)^{\cdot} = 0 \tag{5.141}$$

Exercise 56: Take $T^{\mu\nu}$ to be that of a perfect fluid. Show that the Bianchi identities $\nabla_\nu T^{\mu\nu} = 0$ directly imply (5.141).e.o.e.
The equations we have derived in this section have to be supplemented by the equation of state of the fluid in question. One important choice is that of dust, i.e. $p = 0$, which is believed to be a good approximation to the present universe. Another choice is $p = \frac{\rho}{3}$. This leads to $T_\mu{}^\mu = 0$, and thus corresponds to the averaged version of pure electromagnetic radiation like the one making up the cosmic microwave background.

5.12.4 The spatial geometry

We have not used all the information coming from the Einstein equations. But it would not be very hard to see that the only remaining information can be expressed in the form

$$\mathcal{R}_{ij} = \frac{\mathcal{R}}{3}\, g_{ij}\,, \tag{5.142}$$

and this in fact we already know by virtue of the cosmological principle. Namely, by an argument analogous to that leading to (5.120), $\mathcal{R}_{ij}$ has to be proportional to g_{ij}, and the proportionality factor, by taking the trace, has to be equal to $\frac{\mathcal{R}}{3}$. Furthermore, (5.142) implies that the Riemann tensor of h_{ij} has the form

$$\mathcal{R}_{ijkl} = \frac{\mathcal{R}}{3}\, g_{k[i} g_{j]l} \tag{5.143}$$

The reason is that in 3 dimensions the number of independent components of the Riemann tensor, namely 6, is the same as that of the components of the Ricci tensor, so the former already uniquely determines the latter. Defining a time independent metric by writing $g_{ij}(t,x) = R^2(t)\,\bar{g}_{ij}(x)$ and, using (5.126), Eq.(5.143) takes the form

$$\bar{\mathcal{R}}_{ijkl} = 2k\,\bar{g}_{k[i}\bar{g}_{j]l} \tag{5.144}$$

This relation determines the three types of what is called a space form or a space of constant curvature. It turns out (see the script by Prof. Rumpf for more details) that $\bar{g}_{ij}$ can be written as

$$\bar{g}_{ij}dx^i dx^j = \frac{dr^2}{1-kr^2} + r^2 d\Omega^2 \tag{5.145}$$

The line element (5.145) is for $k = 0$ simply the flat metric, as it has to be.
Exercise 57: Show that (5.145) for $k = -1$ can be realized as the metric on $\mathbb{H}^3$, i.e. the metric induced by the Minkowski metric $\eta_{\mu\nu}$ on $\mathbb{R}^4$ on the three surfaces given by $\eta_{\mu\nu}x^\mu x^\nu = -1$. e.o.e.
Exercise 58: Show that (5.145) for $k = 1$ can be realized as the metric on $\mathbb{S}^3$, i.e. the metric induced by the Euclidean metric $\delta_{\mu\nu}$ on $\mathbb{R}^4$ on the three surface given by $\delta_{\mu\nu}x^\mu x^\nu = 1$. e.o.e.

Chapter 6

Appendices

6.1 A: Covariant differentiation

In section (1.2) we were led to the question of the form of the derivative of a covector field in arbitrary coordinates. A similar question can be asked about the derivative of a vector field or, in fact, any tensor field with the exception of scalar fields, whose derivative ("gradient") always makes sense as a covector field. Consider a covector field ω_i for example, given in standard coordinates on $\mathbb{R}^3$. We require the covariant derivative to be the partial derivative in these coordinates. Using (1.30), the chain and the Leibniz rule, it is clear that the partial derivative in any other coordinate system differs from the covariant derivative, i.e. the partial derivative transformed as if it were a $(0, 2)$ tensor, by a term determined by a three index object $\bar{\Gamma}^i_{jk}$ in the following way

$$\bar{D}_i \bar{\omega}_j = \bar{\partial}_i \bar{\omega}_j - \bar{\Gamma}^k_{ij} \bar{\omega}_k \tag{6.1}$$

Explicitly we have that

$$\bar{\Gamma}^i_{jk}(\bar{x}) = -\frac{\partial x^l}{\partial \bar{x}^j}(\bar{x}) \frac{\partial x^m}{\partial \bar{x}^k}(\bar{x}) \frac{\partial^2 \bar{x}^i}{\partial x^l \partial x^m}(x(\bar{x})) \tag{6.2}$$

For later use note that

$$\Gamma^i_{jk} = \Gamma^i_{kj} \tag{6.3}$$

To write down formulae for the covariant derivative of tensor fields of arbitrary rank, it is useful to have some general rules rather than repeat a derivation analogous to the above. Note, first of all, that the partial derivative in the original coordinates is additive and satisfies the Leibniz rule w.r. to arbitrary tensor products, i.e.

$$D_i(s^{\cdot\cdot}{}_{\cdot\cdot}\, t^{\cdot\cdot}{}_{\cdot\cdot\cdot}) = (D_i s^{\cdot\cdot}{}_{\cdot\cdot})(t^{\cdot\cdot}{}_{\cdot\cdot\cdot}) + (s^{\cdot\cdot}{}_{\cdot\cdot})(D_i t^{\cdot\cdot}{}_{\cdot\cdot\cdot}) \tag{6.4}$$

Using that every, say covariant, tensor is a linear superposition of tensor products of covectors, it follows that, for example for a $(0,2)$ tensor

$$D_i s_{jk} = \partial_i s_{jk} - \Gamma^l_{ij} s_{lk} - \Gamma^l_{ik} s_{jl} \tag{6.5}$$

How about contravariant tensors? Note the identity for partial derivatives

$$\partial_i(\omega_j v^j) = (\partial_i \omega_j) v^i + \omega_j(\partial_i v^j) \tag{6.6}$$

Thus, in arbitrary coordinates, we have to require the same identity for the covariant derivative. Thus

$$D_i(\omega_j v^j) = (D_i \omega_j) v^i + \omega_j(D_i v^j) \tag{6.7}$$

But the covariant derivative on the l.h. side is a partial derivative. Using this and (6.6) we find after one cancellation that

$$\omega_j(\partial_i v^j) = \omega_j D_i v^j - \Gamma^k_{ij} \omega_k v^j \tag{6.8}$$

Since ω_i is arbitrary it follows that

$$D_i v^j = \partial_i v^j + \Gamma^j_{ik} v^k \,. \tag{6.9}$$

There is an analogous formula for the covariant derivative of arbitrary contravariant tensor fields. For mixed tensor fields we have for example

$$D_i t^k{}_j = \partial_i t^k{}_j + \Gamma^k_{il} t^l{}_j - \Gamma^l_{ij} t^k{}_l \tag{6.10}$$

We are now able to replace the inconvenient formula (6.2), as follows: we know that the contravariant Euclidean metric is constant in the original coordinates, i.e. that $\partial_i h^{jk} = \partial_i \delta^{jk} = 0$. We now introduce the covariant metric defined by the equation

$$h^{ij} h_{jk} = 0 \tag{6.11}$$

Since h^{ij}, viewed as a metric, is nonsingular, the metric h_{ij} exists and is unique. It is again symmetric, i.e. $h_{ij} = h_{ji}$. In fact, in the given coordinates, its components are a given by $h_{ij} = \delta_{ij}$. Furthermore we have that

$$\partial_i h_{jk} = 0 \tag{6.12}$$

and consequently, in arbitrary coordinates there holds

$$D_i h_{jk} = 0 \tag{6.13}$$

Using our formulas for the covariant derivative this implies

$$\partial_i h_{jk} = \Gamma^l_{ij} h_{lk} + \Gamma^l_{ik} h_{jl} \tag{6.14}$$

or

$$\partial_k h_{ij} = \Gamma^l_{ki} h_{lj} + \Gamma^l_{kj} h_{il} \tag{6.15}$$

or

$$\partial_j h_{ki} = \Gamma^l_{jk} h_{li} + \Gamma^l_{ji} h_{kl} \,. \tag{6.16}$$

Now adding (6.14) and (6.15) and subtracting (6.16) and using the symmetry (6.3) we finally find

$$2h_{il}\Gamma^l_{jk} = \partial_j h_{ik} + \partial_k h_{ij} - \partial_i h_{jk} \,, \tag{6.17}$$

what is the same as

$$\Gamma^i_{jk} = \frac{1}{2} h^{il} (\partial_j h_{kl} + \partial_k h_{jl} - \partial_l h_{jk}) \tag{6.18}$$

6.2 B: Symmetries of Euclidean space

We are seeking transformations $x^i \mapsto \bar{x}^i(x)$ which are symmetries of the Euclidean metric. In other words there should hold

$$\frac{\partial x^i}{\partial \bar{x}^k} \frac{\partial x^j}{\partial \bar{x}^l} \, \delta_{ij} = \delta_{kl} \tag{6.19}$$

After taking a partial derivative it follows that

$$\left(\frac{\partial^2 x^i}{\partial \bar{x}^k \partial \bar{x}^l} \frac{\partial x^j}{\partial \bar{x}^m} + \frac{\partial x^i}{\partial \bar{x}^k} \frac{\partial^2 x^j}{\partial \bar{x}^m \partial \bar{x}^l} \right) \delta_{ij} = 0 \tag{6.20}$$

By a manoeuvre similar to that leading to Eq.(6.17), we find that each term in (5.39) is separately zero. Using that the Jacobi matrix $\frac{\partial x^j}{\partial \bar{x}^m}$ is nonsingular we end up with

$$\frac{\partial^2 x^i}{\partial \bar{x}^k \partial \bar{x}^l} = 0 \,, \tag{6.21}$$

which has as general solution

$$x^i = M^i{}_j \bar{x}^j + d^i \tag{6.22}$$

where the matrix $M^i{}_j$ and the vector d^i are constant. Inserting back into (6.19) we see that $M^i{}_j$ has to be a rotation matrix, i.e.

$$M^k{}_i M^l{}_j \, \delta_{kl} = \delta_{ij} \tag{6.23}$$

Of course the inverse transformation to (6.22) is again a linear combination of a rotation and a translation.
Suppose we have a 1-parameter family $\bar{x}^i(\epsilon; x) = M^i{}_j(\epsilon) x^j + d^i(\epsilon)$ of Euclidean motions. Defining $N^i{}_j \doteqdot \frac{d}{d\epsilon} M^i{}_j|_{\epsilon=0}$ and $c^i \doteqdot \frac{d}{d\epsilon} c^i|_{\epsilon=0}$, we find that the matrix N is antisymmetric in the sense that

$$2N^k{}_{(i}\, \delta_{j)k} = N_{ji} + N_{ij} = 0 \tag{6.24}$$