General Relativity and Cosmology

Stefan Weinzierl

January 17, 2023

Contents

1 Introduction

These lecture notes are based on a two-semester course on "General Relativity and Cosmology" given at the University of Mainz. The first semester covers the basics of general relativity and black holes, while the second semester is focussed on cosmology and starts with chapter 8.

Text books:

- R. Sexl und H. Urbantke, Gravitation und Kosmologie, Spektrum Akademischer Verlag
- W. Rindler, Relativity, Oxford University Press
- S. Carroll, Spacetime and Geometry, Addison-Wesley
- J. Peacock, Cosmological Physics, Cambridge University Press
- Ch. Misner, K. Thorne and J. Wheeler, Gravitation, Freeman and Company
- S. Weinberg, Gravitation and Cosmology, John Wiley
- G. Ellis and S. Hawking, The Large-Scale Structure of Space-time, Cambridge University Press
- G. Börner, The Early Universe - Facts and Fiction, Springer
- L.D. Landau und E.M. Lifschitz, Band II, Klassische Feldtheorie, Akademie-Verlag
- E. Kolb and M. Turner, The Early Universe, CRC Press
- S. Dodelson, Modern Cosmology, Academic Press
- M. Maggiore, Gravitational Waves, Oxford University Press

Lecture notes:

- D. Hooper, Dark Matter, arXiv:0901.4090
- S. Profumo, Astrophysical Probes of Dark Matter, arXiv:1301.0952
- G. Gelmini, The Hunt for Dark Matter, arXiv:1502.01320
- D. Baumann, Lectures on Inflation, arXiv:0907.5424
- E. Flanagan and S. Hughes, The basics of gravitational wave theory, arXiv:gr-qc/0501041

1.1 History

1638	G. Galilei	Principle of relativity
1676	O. Rømer	speed of light is finite
	Ch. Huygens	1000 earth's diameter per minute
1687	I. Newton	laws of mechanics
1864	J. C. Maxwell	Maxwell's equations
1900	M. Planck	h: Planck's constant
1905	A. Einstein	special relativity
1915	A. Einstein	general relativity
1919	A. Eddingtons	experimental confirmation of general relativity

1.2 Newtonian mechanics

Newton's laws:

1. A free particle moves with constant velocity along straight lines.

2. The force acting on a particle equals the product of its mass and its acceleration:

$$\vec{F} = m\vec{a}.$$

3. The forces of action and reaction have the same absolute value and opposite directions. If particle A exerts a force $\vec{F}$ on particle B, then particle B exerts a force $-\vec{F}$ on particle A.

Remark: Usually we state physical laws with respect to a reference system. A **rigid reference system** is an (imaginary) extension of a rigid body. For example, the earth defines a rigid reference system in the complete space, consisting of all points which are fixed relatively to the earth and among themselves. A concrete example is given by the positions of geostationary satellites.

Among all rigid reference systems the **inertial systems** play a special role. Inertial system are by definition reference systems, where free particles move with constant velocity along straight lines. The inertial systems are the reference systems where Newton's laws are valid.

Remark: Newton postulated the existence of an absolute space, which he identified with the centre-of-mass system of the solar system. In addition, Newton assumed the concept of an absolute time.

Galilei transformations: Given two inertial systems K and K', such that the origin of K moves with velocity v along the x-axis of K', the Galilei transformation reads

$$x' = x + vt, \quad y' = y, \quad z' = z, \quad t' = t.$$

2 Special relativity

2.1 Postulates

Inertial system: Reference system, in which a force-free body moves with constant velocity.

The relative velocity of one inertial system against another inertial system is constant.

Principle of relativity: The law of nature have the same form in all inertial systems.

Principle of a finite signal speed (i.e. there exists a maximal speed of action propagation).

The signal speed has the same value in every inertial system and equals the speed of light

$$c = 2.99792 \cdot 10^8 \mathrm{m/s}.$$

The limit case of classical mechanics: $c \to \infty$. Within classical mechanics we have Galilei's principle of relativity: Spatial relations depend on the reference system. Time is considered as an absolute quantity.

Within special relativity time is no longer an absolute quantity. Example: Consider two inertial systems K and K', where K moves relative to K' along the x'-axis. Assume further that the direction of the x-axis in K coincides with the direction of the x'-axis in K'. Assume now that from a point A on the x-axis one emits a signal in the positive and negative x-direction. Since the signal speed in system K equals c in any direction, the signal will reach two points B and C, which are located at equal distance from A, but in opposite directions, at the same time. However, these two events (arrival of the signal at point B, respectively C) do not occur at the same time for an observer in system K'.

2.2 Distance, metric and four-vectors

An event is characterised by the spatial position, where it takes place and by the time, when it takes place. Thus, an event is characterised by three spatial coordinates and one time coordinate, which together form a four-dimensional space.

Consider again the reference systems K and K': Consider two events: The first event is defined by emitting at the position (x_1, y_1, z_1) at the time t_1 a light signal. This light signal arrives at time t_2 at position (x_2, y_2, z_2), which defines the second event. Since the signal propagates with the speed of light, it has travelled the distance

$$c(t_2 - t_1).$$

On the other hand, the distance is of course also given by

$$\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}.$$

Therefore we have:

$$c^2(t_2-t_1)^2-(x_1-x_2)^2-(y_1-y_2)^2-(z_1-z_2)^2 \;=\; 0.$$

Let us denote in the system K' the coordinates of the first event by x'_1, y'_1, z'_1, t'_1 and the coordinates of the second event by x'_2, y'_2, z'_2, t'_2. Since the speed of light has the same value c in all inertial coordinate systems, we have with the same argumentation as above

$$c^2(t'_2-t'_1)^2-(x'_1-x'_2)^2-(y'_1-y'_2)^2-(z'_1-z'_2)^2 \;=\; 0.$$

Definition: Denote by x_1, y_1, z_1, t_1 and x_2, y_2, z_2, t_2 the coordinates of two arbitrary events. We call the quantity

$$s_{12} \;=\; \sqrt{c^2(t_2-t_1)^2-(x_1-x_2)^2-(y_1-y_2)^2-(z_1-z_2)^2}$$

the **distance** between these two events.

From the invariance of the speed of light it follows that if the distance between two events vanishes in one reference system, it will also vanish in all other reference systems.

More general we have: The distance between two events is the same in all reference systems. Proof: We first consider two events, which are separated by an infinitesimal distance

$$ds^2 \;=\; c^2dt^2-dx^2-dy^2-dz^2.$$

The vanishing of the infinitesimal distance $ds=0$ in one inertial system implies the vanishing of the infinitesimal distance $ds'=0$ in any other system. ds and ds' are infinitesimal quantities of the same order. These two facts imply that they have to be proportional:

$$ds^2 \;=\; a\,ds'^2.$$

The constant of proportionality a cannot depend on space- and time coordinates, as this would contradict the homogeneity of space-time. Furthermore, a cannot depend on the direction of the relative velocity between the two reference systems, as this would contradict the isotropy of space. This implies that a can only depend on the absolute value of the relative velocity between the two inertial systems. Consider now the reference systems K, K_1 and K_2. Let $\vec{v}_1$ be the velocity of K_1 relative to K, let $\vec{v}_2$ be the velocity of K_2 relative to K and let $\vec{v}_{12}$ be the velocity of K_2 relative to K_1. We have

$$ds^2=a(v_1)ds_1^2, \;\; ds^2=a(v_2)ds_2^2, \;\; ds_1^2=a(v_{12})ds_2^2,$$

and therefore

$$\frac{a(v_2)}{a(v_1)} \;=\; a(v_{12}).$$

Since v_{12} depends on the angle between $\vec{v}_1$ and $\vec{v}_2$, so does the right-hand side. However, the left-hand side does not depend on the angle. It follows, that $a(v)$ must be a constant, and from the same equation it follows that the constant must be equal to 1. Therefore

$$ds^2 = ds'^2,$$

and the equality of the infinitesimal distances implies the equality of finite distances:

$$s = s'.$$

$s_{12}^2 > 0$ time-like distance;
there exists a reference systems, where the events 1 and 2 occur at the same spatial position.

$s_{12}^2 < 0$ space-like distance;
there exists a reference systems, where the events 1 and 2 occur at the same time.

$s_{12}^2 = 0$ light-like distance;
light cone

Two events can only be causally connected, if the distance between them satisfies $s_{12} \geq 0$. This follows immediately from the fact, that no casual action can propagate with a speed greater than the speed of light.

Four-vectors: We may view the coordinates (ct, x, y, z) of an event as the components of a vector in a four-dimensional space.

$$x^0 = ct, \quad x^1 = x, \quad x^2 = y, \quad x^3 = z.$$

$$\begin{aligned} x^\mu &= (x^0, x^1, x^2, x^3), \\ &= (x^0, \vec{x}). \end{aligned}$$

We use greek indices $\mu, \nu, ...$, which take the values $0, 1, 2, 3$, to denote the components of a four-vector. Latin indices $i, j, ...$, which take the values $1, 2, 3$, are used to denote the components of a spatial three-vector.

The distance of two evens x_a and x_b is given by

$$s_{ab}^2 = (x_a^0 - x_b^0)^2 - (x_a^1 - x_b^1)^2 - (x_a^2 - x_b^2)^2 - (x_a^3 - x_b^3)^2.$$

We define the metric tensor $g_{\mu\nu}$ by

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

This allows us to write the distance as

$$s_{ab}^2 \;=\; \sum_{\mu=0}^{3}\sum_{\nu=0}^{3} g_{\mu\nu}\left(x_a^\mu - x_b^\mu\right)\left(x_a^\nu - x_b^\nu\right).$$

Einstein's summation convention: Sums as above are often written without the summation sign. In general, Einstein's summation convention is the rule, that indices which occur in pairs imply a summation over all values of this index. The summation sign is not written explicitly. For each pair of indices, one index must occur as a subscript, the other as a superscript.
Therefore

$$s_{ab}^2 \;=\; g_{\mu\nu}\left(x_a - x_b\right)^\mu \left(x_a - x_b\right)^\nu.$$

We call a four-vector x^μ with an upper index a contravariant four-vector, a four-vector x_μ with a lower index is called a covariant four-vector. The relation between covariant and contravariant four-vectors is given by

$$x_\mu \;=\; g_{\mu\nu}x^\nu, \quad x^\mu \;=\; g^{\mu\nu}x_\nu, \quad g^{\mu\nu} \;=\; \left(g^{-1}\right)^{\mu\nu} \;=\; \mathrm{diag}(1,-1,-1,-1).$$

Thus, we may write the distance equally well as

$$s_{ab}^2 \;=\; \left(x_a - x_b\right)_\mu \left(x_a - x_b\right)^\mu = \left(x_a - x_b\right)^\mu \left(x_a - x_b\right)_\mu.$$

Remark: The geometry defined by the quadratic form $g_{\mu\nu} = \mathrm{diag}(1,-1,-1,-1)$ is not an Euclidean geometry. One speaks of a pseudo-Euclidean geometry. The special case of a four-dimensional space with the metric $\mathrm{diag}(1,-1,-1,-1)$ is known as Minkowski space.

2.3 Proper time

Consider the following situation: We observe from an inertial system K' a moving clock. The motion of the clock may be arbitrary. We may approximate the motion of the clock by sequence of motions with constant velocity. Thus, we may associate for every time t an inertial system K to the clock, such that the clock is at rest in K at time t. (If the clock is accelerating, we will need different inertial systems at different times.) In the original system K' the clock travels in the infinitesimal time interval dt' the spatial distance

$$\sqrt{dx'^2 + dy'^2 + dz'^2}.$$

We may ask, what time the clock displays in system K at the end of this infinitesimal trajectory. Phrased differently, we ask what is the infinitesimal time interval dt in K. From the invariance of the distance we have

$$c^2 dt'^2 - dx'^2 - dy'^2 - dz'^2 \;=\; c^2 dt^2$$

and therefore

$$dt = dt'\sqrt{1-\frac{dx'^2+dy'^2+dz'^2}{c^2dt'^2}} = dt'\sqrt{1-\frac{v^2}{c^2}}.$$

Integration gives for an arbitrary motion

$$t_2 - t_1 = \int\limits_{t_1'}^{t_2'} dt'\sqrt{1-\frac{v^2}{c^2}}.$$

With $t_1 = t_1' = 0$ this simplifies to

$$t_2 = \int\limits_{0}^{t_2'} dt'\sqrt{1-\frac{v^2}{c^2}}.$$

t_2 is called the **proper time** of the moving object.

Remark 1: The proper time of a moving object is always smaller than the corresponding time interval in a non-moving reference system.

Remark 2: This is no contradiction to the principle of relativity, since for a comparison of the clocks we need one clock in the moving system but several clocks in the non-moving system.

Remark 3: Also a clock, whose spatial motion is given by a closed curve, does not contradict the principle of relativity. Such a clock cannot be at rest in a single inertial systems at all times.

2.4 Lorentz transformations

Let K and K' be two inertial systems. We would like to have a formula which allows to compute the coordinates x', y', z', t' of an event in the inertial system K', given that we know the coordinates x, y, z, t of the same event in system K.

Recall: The Galilei transformation:

$$x' = x + vt, \quad y' = y, \quad z' = z, \quad t' = t.$$

System K moves with velocity v relative to system K' along the x-axis.

The relativistic generalisation has to keep the distance invariant. This implies that we only have to consider translations and rotations. Translations correspond to a redefinition of the origin of the coordinate system and are not new. We therefore focus on rotations. Each rotation in the four-dimensional space can be decomposed into the six basic rotations in the planes xy, yz, zx, tx, ty and tz. Basic rotations in the first three planes (xy, yz and zx) correspond to ordinary spatial rotations. Let us therefore consider as an example a rotation in the tx-plane. This leaves the y- and z-coordinates unchanged. The rotation has to keep the difference

$$ct^2 - x^2$$

invariant. Due to the pseudo-Euclidean metric with a minus sign we either obtain an imaginary rotation angle or (converting sin and cos with imaginary arguments to sinh and cosh) hyperbolic functions:

$$\begin{aligned} ct' &= x \sinh\phi + ct \cosh\phi, \\ x' &= x \cosh\phi + ct \sinh\phi, \end{aligned}$$

or in four-vector notation

$$x'^{\mu} = \Lambda^{\mu}{}_{\nu} x^{\nu},$$

with

$$\Lambda^{\mu}{}_{\nu} = \begin{pmatrix} \cosh\phi & \sinh\phi & 0 & 0 \\ \sinh\phi & \cosh\phi & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Determination of ϕ: We consider the origin of the system K in K':

$$ct' = ct \cosh\phi, \qquad x' = ct \sinh\phi,$$

therefore

$$\tanh\phi = \frac{x'}{ct'} = \frac{v}{c}.$$

Thus

$$\sinh\phi = \frac{\frac{v}{c}}{\sqrt{1-\frac{v^2}{c^2}}}, \qquad \cosh\phi = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}}.$$

In the limit $v \ll c$ we recover the Galilei transformation.

Common abbreviations:

$$\beta = \frac{v}{c}, \qquad \gamma = \frac{1}{\sqrt{1-\frac{v^2}{c^2}}}.$$

Length contraction: A rod of length l, which is at rest in system K and oriented parallel to the x-axis, has in system K' the length

$$l' = l\sqrt{1-\frac{v^2}{c^2}}.$$

(In order to prove this formula determine the x'-coordinates x'_1 and x'_2 of the two end points of the rod at a common time t' in system K'.)

2.5 Transformation of the velocity

Assume that system K moves relative to system K' with the velocity V in the direction of the positive x-axis. Let the velocity of a particle in system K be

$$v_x = \frac{dx}{dt}, \quad v_y = \frac{dy}{dt}, \quad v_z = \frac{dz}{dt},$$

and denote the corresponding velocity in system K' by

$$v'_x = \frac{dx'}{dt'}, \quad v'_y = \frac{dy'}{dt'}, \quad v'_z = \frac{dz'}{dt'}.$$

The infinitesimal quantities are related by the Lorentz transformation

$$dx' = \gamma(dx + V dt), \quad dy' = dy, \quad dz' = dz, \quad dt' = \gamma\left(dt + \frac{V}{c^2}dx\right).$$

Division of the first three equations by the fourth equation gives:

$$v'_x = \frac{v_x + V}{1 + \frac{v_x V}{c^2}}, \quad v'_y = \frac{v_y}{\gamma(1 + \frac{v_x V}{c^2})}, \quad v_z = \frac{v_z}{\gamma(1 + \frac{v_x V}{c^2})}.$$

Special case: $v_x = v$, $v_y = v_z = 0$:

$$v' \quad = \quad \frac{v + V}{1 + \frac{vV}{c^2}}.$$

If we calculate v' with the help of this formula, the result will always be smaller or equal than c.

2.6 The four-velocity

The four-velocity of a particle is the four-vector

$$u^\mu \quad = \quad \frac{dx^\mu}{ds},$$

where ds is the infinitesimal proper time interval in units of length. Explicitly, ds is given by

$$ds \quad = \quad cdt\sqrt{1 - \frac{v^2}{c^2}},$$

where v is the usual (spatial) speed of the particle. Therefore

$$u^1 \quad = \quad \frac{dx^1}{ds} = \frac{dx^1}{cdt\sqrt{1 - \frac{v^2}{c^2}}} = \frac{v_x}{c\sqrt{1 - \frac{v^2}{c^2}}}.$$

Repeating this for all components we find

$$u^{\mu} = \left(\frac{1}{\sqrt{1-\frac{v^2}{c^2}}}, \frac{\vec{v}}{c\sqrt{1-\frac{v^2}{c^2}}} \right).$$

The components of u^{μ} are not independent, but satisfy the relation

$$u^{\mu} u_{\mu} = 1.$$

We may interpret the four-velocity geometrically as a unit four-vector, tangent to the world line of the particle.

2.7 The Lorentz group

Group axioms: Let G be a non-empty set with a composition. G is a group, if the following conditions are satisfied:

- Associative law: $a \cdot (b \cdot c) = (a \cdot b) \cdot c$.
- Existence of a neutral element e: $e \cdot a = a$.
- Existence of an inverse element a^{-1} for each element a: $a^{-1} \cdot a = e$.

Example: Matrix groups.

- $GL(n,\mathbb{R})$, $GL(n,\mathbb{C})$: Group of invertible $n \times n$ matrices: $\det M \neq 1$
- $SL(n,\mathbb{R})$, $SL(n,\mathbb{C})$: $\det M = 1$;
- $O(n)$: $MM^T = 1$
- $SO(n)$: $MM^T = 1$ and $\det M = 1$.
- $U(n)$: $MM^{\dagger} = 1$.
- $SU(n)$: $MM^{\dagger} = 1$ and $\det M = 1$.

Definition of the Lorentz group:
Matrix group, which leaves the metric tensor $g_{\mu\nu} = \mathrm{diag}(1,-1,-1,-1)$ invariant:

$$\Lambda^T g \Lambda = g,$$

or equivalently in greater detail with indices:

$$\Lambda^{\mu}{}_{\sigma} g_{\mu\nu} \Lambda^{\nu}{}_{\tau} = g_{\sigma\tau}.$$

This group is denoted by $O(1,3)$. It is easy to see that

$$(\det \Lambda)^2 = 1,$$

and hence

$$\det \Lambda = \pm 1.$$

If in addition $\det \Lambda = 1$ holds, we call this group the **proper Lorentz group** and denote it by $SO(1,3)$.
A further distinction can be made depending on whether the time direction is conserved or not. If

$$\Lambda^0{}_0 \geq 1,$$

the time direction is conserved and we call the corresponding group the **orthochronous Lorentz group**. If on the other hand we have

$$\Lambda^0{}_0 \leq -1,$$

then the time direction is reversed.
Remark:

$$\left|\Lambda^0{}_0\right| \geq 1$$

follows from $\Lambda^\mu{}_\sigma g_{\mu\nu} \Lambda^\nu{}_\tau = g_{\sigma\tau}$ for $\sigma = \tau = 0$:

$$\left(\Lambda^0{}_0\right)^2 - \sum_{j=1}^{3} \left(\Lambda^j{}_0\right)^2 = 1.$$

In summary we find that the Lorentz group consists of four connected components. The connected components are characterised by the values

$$\det \Lambda \quad \text{and} \quad \Lambda^0{}_0.$$

Among the four connected components the **proper orthochronous Lorentz group** defined by

$$\Lambda^\mu{}_\sigma g_{\mu\nu} \Lambda^\nu{}_\tau = g_{\sigma\tau}, \quad \det \Lambda = 1, \quad \Lambda^0{}_0 \geq 1,$$

is of particular interest. (The other three connected components are not groups, as they do not contain the neutral element.) We may obtain the other three connected components from the composition of an element of the proper orthochronous Lorentz group and the two discrete transformations of time reversal

$$\Lambda^\mu{}_\nu = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

and space inversion

$$\Lambda^{\mu}{}_{\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}.$$

The **Poincaré group**: The Poincaré group consists of all elements of the Lorentz group and the translations. The coordinates transform according to

$$x'^{\mu} = \Lambda^{\mu}{}_{\nu} x^{\nu} + b^{\mu}.$$

2.8 Tensors in Minkowski space

Let V be a vector space and G a group. We say G acts on V, if there is a map

$$G \times V \to V$$

such that

$$g_1 (g_2 v) = (g_1 g_2) v.$$

In this case we call V a representation of G.

Example 1: Let V be a n-dimensional vector space and $G = GL(n, \mathbb{R})$. The map $G \times V \to V$ is defined as the multiplication of a matrix with a column vector:

$$v'_i = \sum_{j=1}^{n} M_{ij} v_j$$

example 2: Take V to be Minkowski space and G the Lorentz group.

$$x'^{\mu} = \Lambda^{\mu}{}_{\nu} x^{\nu}, \quad \text{(Einstein's summation convention)}$$

Example 3: Let V be a n^2-dimensional vector space and $G = GL(n, \mathbb{R})$. We write elements of V as v_{ij} with $1 \leq i, j \leq n$. G acts on V as follows:

$$v'_{ij} = \sum_{k=1}^{n} \sum_{l=1}^{n} M_{ik} M_{jl} v_{kl}$$

We call v_{ij} a rank 2 tensor.

Example 4: Let V be a 16-dimensional vector space and G the Lorentz group.

$$T'^{\mu\nu} = \Lambda^{\mu}{}_{\rho} \Lambda^{\nu}{}_{\sigma} T^{\rho\sigma}$$

$T^{\mu\nu}$ is a rank 2 tensor.

Example 5: Let V be a 64-dimensional vector space and G the Lorentz group.

$$T'^{\mu\nu\rho} = \Lambda^{\mu}{}_{\sigma}\Lambda^{\nu}{}_{\kappa}\Lambda^{\rho}{}_{\lambda}T^{\sigma\kappa\lambda}$$

$T^{\mu\nu\rho}$ is a rank 4 tensor.

Let us now give the general definition: Consider a vector space endowed with a group action. A **tensor** is an element of this vector space. The **rank of the tensor** is the number of copies of the group element required to define the group action.

Let us now specialise to Minkowski space and the Lorentz group. We also define pseudotensors. **Pseudotensors** transform as tensors under all transformations of the proper orthochronous Lorentz group. However, the transformation law of a pseudotensor differs by a minus sign from the transformation law of a tensor for the two discrete transformations of time reversal and spatial inversion.

We call a rank 0 pseudotensor a **pseudoscalar** and we call a rank 1 pseudotensor an **axial vector**.

Within special relativity we distinguish in addition between upper and lower indices (contravariant and covariant components). The relation between upper and lower indices is again provided by the metric tensor:

$$T^{\mu}_{\nu} = g_{\nu\rho}T^{\mu\rho}, \qquad T_{\mu\nu} = g_{\mu\rho}g_{\nu\sigma}T^{\rho\sigma}$$

Tensors with particular symmetry properties: A tensor is called **symmetric** in two indices μ and ν, if

$$S^{...\mu...\nu...} = S^{...\nu...\mu...}.$$

A tensor is called **anti-symmetric** in two indices μ and ν, if

$$A^{...\mu...\nu...} = -A^{...\nu...\mu...}.$$

In particular we have for an anti-symmetric rank 2 tensor $A^{00} = A^{11} = A^{22} = A^{33} = 0$.

Examples of tensors appearing within special relativity:

Rank 1: Position vector x^{μ}, momentum vector p^{μ}.

Rank 2: Metric tensor $g^{\mu\nu}$.

Rank 4: Total anti-symmetric tensor (Levi-Civita tensor) $\varepsilon^{\mu\nu\rho\sigma}$. The total anti-symmetric tensor is defined by

$$\begin{aligned} \varepsilon_{0123} &= 1, \\ \varepsilon_{\mu\nu\rho\sigma} &= 1 \quad \text{if } (\mu,\nu,\rho,\sigma \text{ is an even permutation of } (0,1,2,3), \\ \varepsilon_{\mu\nu\rho\sigma} &= -1 \quad \text{if } (\mu,\nu,\rho,\sigma \text{ is an odd permutation of } (0,1,2,3), \\ \varepsilon_{\mu\nu\rho\sigma} &= 0 \quad \text{otherwise.} \end{aligned}$$

The total anti-symmetric tensor is a pseudotensor, the components remain unchanged under time reversal and spatial inversion.

Dual tensors: Let $F^{\mu\nu}$ be an anti-symmetric rank 2 tensor. The pseudotensor

$$\tilde{F}^{\mu\nu} = \frac{1}{2}\varepsilon^{\mu\nu\rho\sigma}F_{\rho\sigma}$$

is called the dual tensor of $F^{\mu\nu}$.

A similar concept applies to vectors A^{μ}: The rank 3 tensor

$$\tilde{A}^{\mu\nu\rho} = \varepsilon^{\mu\nu\rho\sigma}A_{\sigma}$$

is called the dual tensor of A^{μ}.

2.9 Relativistic mechanic

The essential elements of classical mechanics: Within the **Lagrange formalism** one considers generalised coordinates $q_i(t)$ and the corresponding generalised velocities $\dot{q}_i(t) = \frac{\partial}{\partial t}q_i(t)$.
Lagrange function:

$$L(q_i, \dot{q}_i)$$

Action:

$$S[q_i(t)] = \int_{t_a}^{t_b} dt\, L(q_i, \dot{q}_i)$$

Principle of least action: A particle moves in such a way that the action is extremal.

Action for a free matter particle:
- has to be invariant under Lorentz transformations,
- must only contain first order differentials.

This implies that the action for a free particle is of the form

$$S = -\alpha \int_a^b ds.$$

The path of integration is along the worldline of the particle between two events a and b. In order to have a minimum for the action S, we must require $\alpha > 0$. In order to see this, we first consider a particle at rest, for which $ds = cdt$. Let us then consider a trajectory where the particle is moving. We write

$$S = \int_{t_a}^{t_b} L\, dt,$$

where L is called the Lagrange function. With

$$ds = cdt\sqrt{1-\frac{v^2}{c^2}}$$

we obtain

$$L = -\alpha c\sqrt{1-\frac{v^2}{c^2}}.$$

We would like to have that the trajectory where the particle is at rest, is a minimum of the action. Since $\sqrt{1-\frac{v^2}{c^2}} \leq 1$ it follows that we must require $\alpha > 0$. Let us now consider the classical limit:

$$\lim_{c\to\infty} L = \text{const} + \frac{1}{2}mv^2.$$

We expand L in v/c:

$$L = -\alpha c\sqrt{1-\frac{v^2}{c^2}} \approx -\alpha c + \frac{\alpha v^2}{2c}$$

Therefore $\alpha = mc$ and

$$S = -mc^2\int_{t_a}^{t_b} dt\sqrt{1-\frac{v^2}{c^2}}, \qquad L = -mc^2\sqrt{1-\frac{v^2}{c^2}}.$$

The three-momentum of a particle is the vector

$$\vec{p} = \frac{\partial L}{\partial \vec{v}} = \frac{m\vec{v}}{\sqrt{1-\frac{v^2}{c^2}}} \quad \text{with } \vec{v} = \dot{\vec{x}}, \quad \left(\text{recall}: \; p_i = \frac{\partial L}{\partial \dot{q}_i}\right).$$

The energy of a particle is the quantity

$$\begin{aligned} E &= \vec{p}\vec{v} - L, \qquad (\text{recall}: \; E = p_i\dot{q}_i - L), \\ &= \frac{m\vec{v}}{\sqrt{1-\frac{v^2}{c^2}}}\vec{v} + mc^2\sqrt{1-\frac{v^2}{c^2}} = \frac{m}{\sqrt{1-\frac{v^2}{c^2}}}\left(v^2+c^2-v^2\right) = \frac{mc^2}{\sqrt{1-\frac{v^2}{c^2}}}, \end{aligned}$$

For small velocities we obtain

$$E \approx mc^2 + \frac{1}{2}mv^2.$$

mc^2 is called the **rest energy**.

Derivation of the equation of motion for a free particle in four-vector notation: We start from

$$S = -mc\int_a^b ds.$$

Variation of the coordinates:

$$x^\mu \rightarrow x^\mu + \delta x^\mu$$

Principle of variation:

$$\frac{\delta}{\delta x^\mu(t)} S[x^\mu(t)] = 0.$$

Auxiliary calculation:

$$\frac{\delta}{\delta x^\mu} ds = \frac{\delta}{\delta x^\mu}\sqrt{dx_\nu dx^\nu} = \frac{1}{2\sqrt{ds^2}} 2dx_\nu \frac{\delta}{\delta x^\mu} dx^\nu = u_\nu \frac{\delta}{\delta x^\mu} dx^\nu$$

Therefore

$$\delta ds = u_\nu \delta dx^\nu$$

Further

$$\begin{aligned}\delta S &= -mc\int_a^b \delta ds = -mc\int_a^b u_\nu \delta dx^\nu = -mc\int_a^b u_\nu \frac{d\delta x^\nu}{ds} ds \\ &= -mc\, u_\nu \delta x^\nu \big|_a^b + mc\int_a^b \left(\frac{d}{ds}u_\nu\right)\delta x^\nu ds\end{aligned}$$

We therefore have

$$\frac{d}{ds}u_\nu = 0,$$

i.e. the free motion of a particle is a motion with constant four-velocity.

Definition of the contravariant momentum four-vector:

$$p^\mu = (E/c, \vec{p}) = \left(\frac{mc}{\sqrt{1-\frac{v^2}{c^2}}}, \frac{m\vec{v}}{\sqrt{1-\frac{v^2}{c^2}}}\right) = mcu^\mu$$

Remark: p^2 is Lorentz invariant.

3 Electrodynamics

3.1 Maxwell's equations

Maxwell's equations:

$$\begin{aligned}
\vec{\nabla}\cdot\vec{B}(t,\vec{x}) &= 0,\\
\vec{\nabla}\times\vec{E}(t,\vec{x})+\frac{1}{c}\frac{\partial}{\partial t}\vec{B}(t,\vec{x}) &= 0,\\
\vec{\nabla}\cdot\vec{E}(t,\vec{x}) &= 4\pi\rho(t,\vec{x}),\\
\vec{\nabla}\times\vec{B}(t,\vec{x})-\frac{1}{c}\frac{\partial}{\partial t}\vec{E}(t,\vec{x}) &= \frac{4\pi}{c}\vec{j}(t,\vec{x}).
\end{aligned}$$

Potentials:

$$\begin{aligned}
\vec{E}(t,\vec{x}) &= -\vec{\nabla}\Phi(t,\vec{x})-\frac{1}{c}\frac{\partial}{\partial t}\vec{A}(t,\vec{x}),\\
\vec{B}(t,\vec{x}) &= \vec{\nabla}\times\vec{A}(t,\vec{x}).
\end{aligned}$$

Gauge transformation:

$$\begin{aligned}
\Phi'(t,\vec{x}) &= \Phi(t,\vec{x})-\frac{1}{c}\frac{\partial}{\partial t}\chi(t,\vec{x}),\\
\vec{A}'(t,\vec{x}) &= \vec{A}(t,\vec{x})+\vec{\nabla}\chi(t,\vec{x}),
\end{aligned}$$

Lorentz force:

$$\vec{F}(t,\vec{x}) = q\left(\vec{E}(t,\vec{x})+\frac{\vec{v}}{c}\times\vec{B}(t,\vec{x})\right).$$

Equivalently we may present electrodynamics in a manifest covariant form. We recall the formula for the four-velocity

$$u^\mu = \frac{dx^\mu}{ds} = \left(\frac{1}{\sqrt{1-\frac{v^2}{c^2}}},\frac{\vec{v}}{c\sqrt{1-\frac{v^2}{c^2}}}\right) = \gamma\left(1,\frac{1}{c}\vec{v}\right).$$

$$ds = \frac{c}{\gamma}dt,$$

We introduce the four-acceleration:

$$w^\mu = \frac{du^\mu}{ds}$$

The relativistic generalisation of Newton's law $\vec{F}=m\vec{a}$:

$$mc^2\frac{d}{ds}u^\mu = K^\mu.$$

Contraction with u_μ gives:

$$mc^2 u_\mu \frac{d}{ds} u^\mu \;=\; \frac{1}{2} mc^2 \frac{d}{ds} \underbrace{u^2}_{1} = 0,$$

and therefore

$$K^\mu u_\mu \;=\; 0.$$

For the spatial components we have

$$\vec{K} \;=\; \gamma \vec{F}$$

We apply this to the Lorentz force:

$$mc^2 \frac{d}{ds} \vec{u} \;=\; mc^2 \frac{d}{ds} \left(\gamma \frac{\vec{v}}{c} \right) = q\gamma \left(\vec{E} + \frac{\vec{v}}{c} \times \vec{B} \right).$$

For the time component we have

$$\begin{aligned} u_\mu K^\mu &= \gamma K^0 - \frac{\gamma}{c} \vec{v}\vec{K} = 0, \\ K^0 &= \frac{1}{c} \vec{v}\vec{K}, \end{aligned}$$

and therefore

$$mc^2 \frac{d}{ds} u^0 \;=\; mc^2 \frac{d}{ds} \gamma = \frac{1}{c} q\gamma \vec{E}\vec{v}.$$

In summary we have:

$$\begin{aligned} mc^2 \frac{d}{ds} (\gamma) &= \gamma q \vec{E} \cdot \frac{\vec{v}}{c}, \\ mc^2 \frac{d}{ds} \left(\gamma \frac{\vec{v}}{c} \right) &= \gamma q \left(\vec{E} + \frac{\vec{v}}{c} \times \vec{B} \right). \end{aligned}$$

The left-hand side may be written covariantly as

$$mc^2 \frac{d}{ds} u^\mu$$

Let us now set

$$F^{\mu\nu} \;=\; \begin{pmatrix} 0 & -E^x & -E^y & -E^z \\ E^x & 0 & -B^z & B^y \\ E^y & B^z & 0 & -B^x \\ E^z & -B^y & B^x & 0 \end{pmatrix}.$$

With this definition we have

$$
\begin{aligned}
F^{\mu\nu} u_\nu &= \begin{pmatrix} 0 & -E^x & -E^y & -E^z \\ E^x & 0 & -B^z & B^y \\ E^y & B^z & 0 & -B^x \\ E^z & -B^y & B^x & 0 \end{pmatrix} \begin{pmatrix} \gamma \\ -\gamma\frac{v^x}{c} \\ -\gamma\frac{v^y}{c} \\ -\gamma\frac{v^z}{c} \end{pmatrix} \\
&= \begin{pmatrix} \gamma \vec{E}\frac{\vec{v}}{c} \\ \gamma E^x + \frac{\gamma}{c}\left(v^y B^z - v^z B^y\right) \\ \gamma E^y + \frac{\gamma}{c}\left(v^z B^x - v^x B^z\right) \\ \gamma E^z + \frac{\gamma}{c}\left(v^x B^y - v^y B^x\right) \end{pmatrix} = \gamma \begin{pmatrix} \vec{E}\frac{\vec{v}}{c} \\ \vec{E} + \frac{\vec{v}}{c} \times \vec{B} \end{pmatrix}.
\end{aligned}
$$

Thus we arrive at

$$
mc^2 \frac{d}{ds} u^\mu = q F^{\mu\nu} u_\nu
$$

The left-hand side transforms as a contravariant four-vector under Lorentz transformations, u_ν transforms as a covariant four-vector. This implies that $F^{\mu\nu}$ must transform as a contravariant rank 2 tensor:

$$
F'^{\mu\nu} = \Lambda^\mu{}_\rho \Lambda^\nu{}_\sigma F^{\rho\sigma}
$$

We call $F^{\mu\nu}$ the **field strength tensor**. We obtain the electric and the magnetic field from $F^{\mu\nu}$ through

$$
\begin{aligned}
E^i &= F^{i\,0} = -F^{0\,i}, \\
B^i &= -\frac{1}{2} \sum_{j,k=1}^{3} \varepsilon_{ijk} F^{jk}.
\end{aligned}
$$

Remark: $F^{\mu\nu}$ is anti-symmetric:

$$
F^{\mu\nu} = -F^{\nu\mu}.
$$

Summary of the covariant formulation:

Definition of the field strength tensor:

$$
F^{\mu\nu} = \begin{pmatrix} 0 & -E^x & -E^y & -E^z \\ E^x & 0 & -B^z & B^y \\ E^y & B^z & 0 & -B^x \\ E^z & -B^y & B^x & 0 \end{pmatrix}
$$

Maxwell's equations:

$$
\begin{aligned}
\partial^\lambda F^{\mu\nu} + \partial^\mu F^{\nu\lambda} + \partial^\nu F^{\lambda\mu} &= 0, \\
\partial_\mu F^{\mu\nu} &= \frac{4\pi}{c} j^\nu,
\end{aligned}
$$

with $j^\mu = (c\rho, \vec{j})$.
Remark: With the help of the total anti-symmetric tensor $\varepsilon_{\mu\nu\rho\sigma}$ and due to the anti-symmetry of $F^{\mu\nu}$ we may rewrite the first equation as

$$\varepsilon_{\mu\nu\rho\sigma}\partial^\nu F^{\rho\sigma} = 0.$$

Lorentz force:

$$mc^2 \frac{d}{ds} u^\mu = q F^{\mu\nu} u_\nu$$

Four-potential:

$$A^\mu = \left(\Phi, \vec{A}\right),$$

$$F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu.$$

Inhomogeneous Maxell's equation:

$$\Box A^\nu - \partial^\nu \partial_\mu A^\mu = \frac{4\pi}{c} j^\nu.$$

Lorenz gauge:

$$\partial_\mu A^\mu = 0$$

Inhomogeneous Maxell's equation in Lorenz gauge:

$$\Box A^\nu = \frac{4\pi}{c} j^\nu.$$

3.2 Lagrange density for the interaction of a particle with the electromagnetic field

Recall: Action for a free particle:

$$S_{\text{particle}} = -mc \int\limits_a^b ds$$

For the interaction between a particle and the electromagnetic field we set

$$S_{\text{interaction}} = -\frac{q}{c} \int\limits_a^b dx^\mu \, A_\mu(x).$$

For a particle we have to consider

$$S_{\text{particle}} + S_{\text{interaction}} = -mc\int_a^b ds - \frac{q}{c}\int_a^b dx^\mu\, A_\mu(x).$$

Variation of the coordinates:

$$x^\mu \;\to\; x^\mu + \delta x^\mu$$

Principle of variation:

$$\delta\left(S_{\text{particle}} + S_{\text{interaction}}\right) = 0.$$

Recall

$$\delta ds = u_\nu d\delta x^\nu.$$

Furthermore

$$\delta\left(A_\mu dx^\mu\right) = A_\mu d\delta x^\mu + \left(\delta A_\mu\right) dx^\mu$$

and

$$\delta A_\mu = A_\mu(x+\delta x) - A_\mu(x) = \left(\partial_\nu A_\mu\right)\delta x^\nu.$$

Hence

$$\begin{aligned}
\delta\left(S_{\text{particle}} + S_{\text{interaction}}\right) &= -mc\int_a^b \delta ds - \frac{q}{c}\int_a^b \delta(dx^\mu\, A_\mu(x)) \\
&= -mc\int_a^b u_\nu d\delta x^\nu - \frac{q}{c}\int_a^b A_\mu d\delta x^\mu - \frac{q}{c}\int_a^b \left(\delta A_\mu\right) dx^\mu \\
&= mc\int_a^b \left(\frac{d}{ds}u_\nu\right)\delta x^\nu ds - \frac{q}{c}\int_a^b A_\mu d\delta x^\mu - \frac{q}{c}\int_a^b \left(\partial_\nu A_\mu\right) dx^\mu \delta x^\nu
\end{aligned}$$

We also have

$$\int_a^b \left(\partial_\nu A_\mu\right) dx^\mu \delta x^\nu = \int_a^b \left(\partial_\nu A_\mu\right) u^\mu \delta x^\nu ds,$$

$$\begin{aligned}
\int_a^b A_\mu d\delta x^\mu &= \int_a^b A_\mu \frac{d}{ds}\delta x^\mu ds = -\int_a^b \left(\frac{d}{ds}A_\mu\right)\delta x^\mu ds = -\int_a^b \frac{\partial A_\mu}{\partial x^\nu}\frac{\partial x^\nu}{ds}\delta x^\mu ds \\
&= -\int_a^b \partial_\nu A_\mu u^\nu \delta x^\mu ds = -\int_a^b \partial_\mu A_\nu u^\mu \delta x^\nu ds
\end{aligned}$$

and therefore

$$\delta\left(S_{\text{particle}}+S_{\text{interaction}}\right) = \int_a^b \left(mc\frac{d}{ds}u_\nu + \frac{q}{c}\partial_\mu A_\nu u^\mu - \frac{q}{c}\partial_\nu A_\mu u^\mu\right)\delta x^\nu ds$$
$$= \int_a^b \left(mc\frac{d}{ds}u_\nu + \frac{q}{c}F_{\mu\nu}u^\mu\right)\delta x^\nu ds.$$

It follows that we must have

$$mc\frac{d}{ds}u_\nu + \frac{q}{c}F_{\mu\nu}u^\mu = 0,$$

and therefore

$$mc^2\frac{d}{ds}u_\mu = qF_{\mu\nu}u^\nu.$$

3.3 Lagrange density of electrodynamics

We make the ansatz that the action of electrodynamics consists of a term describing free fields and a term describing the interaction of the fields with matter.

$$S = S_{\text{fields}} + S_{\text{interaction}}$$

In order to construct $S_{\text{interaction}}$ we generalise the expression of the interaction term for a point source towards a general charge density:

$$S_{\text{interaction,point source}} = -\frac{q}{c}\int_a^b dx^\mu\, A_\mu(x).$$

The charge density and the current density of a point source with trajectory $\vec{x}'(t)$ read:

$$\rho(t,\vec{x}) = q\delta^3(\vec{x}-\vec{x}'(t)),$$
$$\vec{j}(t,\vec{x}) = q\vec{v}(t)\delta^3(\vec{x}-\vec{x}'(t)).$$

Therefore

$$j^\mu(x) = \left(c\rho,\vec{j}\right) = qc\int ds\, u^\mu\, \delta^4\left(x-x'(s)\right)$$

and

$$S_{\text{interaction}} = -\sum_i \frac{q_i}{c}\int_a^b dx^\mu\, A_\mu(x)$$
$$\rightarrow -\frac{1}{c^2}\int d^3x\, c\rho(x)\int ds\frac{dx^\mu}{ds}\, A_\mu(x) = -\frac{1}{c^2}\int d^4x \underbrace{c\rho(x)\frac{dx^\mu}{ds}}_{j^\mu(x)} A_\mu(x)$$
$$= -\frac{1}{c^2}\int d^4x\, j^\mu(x)\, A_\mu(x).$$

Let us now turn to the free field part. For the construction of S_{fields} we require:

- Lorentz invariance.
- Superposition principle, i.e. the field equations should be linear differential equations. This implies that the integrand of S_{fields} has to be no higher than quadratic in the field components.
- Physically unique, i.e. gauge invariant. This translates to the requirement that the integrand should be expressed in terms of $F_{\mu\nu}$ and not A_μ.

The simplest ansatz is given by

$$S_{\text{fields}} \;=\; -\frac{1}{16\pi c}\int d^4x F_{\mu\nu}F^{\mu\nu}.$$

Let us therefore consider

$$S_{\text{fields}}+S_{\text{interaction}} \;=\; -\frac{1}{16\pi c}\int d^4x F_{\mu\nu}(x)F^{\mu\nu}(x)-\frac{1}{c^2}\int d^4x\, j^\mu(x)A_\mu(x).$$

With $F_{\mu\nu}=\partial_\mu A_\nu-\partial_\nu A_\mu$ we obtain

$$S_{\text{fields}}+S_{\text{interaction}} \;=\; \int d^4x\left[-\frac{1}{8\pi c}\left(\partial_\mu A_\nu\right)\left(\partial^\mu A^\nu\right)+\frac{1}{8\pi c}\left(\partial_\mu A_\nu\right)\left(\partial^\nu A^\mu\right)-\frac{1}{c^2}j^\mu(x)A_\mu(x)\right].$$

The Lagrange density reads

$$\mathscr{L} \;=\; -\frac{1}{8\pi}\left(\partial_\mu A_\nu\right)\left(\partial^\mu A^\nu\right)+\frac{1}{8\pi}\left(\partial_\mu A_\nu\right)\left(\partial^\nu A^\mu\right)-\frac{1}{c}j^\mu(x)A_\mu(x).$$

The Euler-Lagrange equations read

$$\frac{\partial\mathscr{L}}{\partial A_\nu}-\partial_\mu\frac{\partial\mathscr{L}}{\partial\left(\partial_\mu A_\nu\right)} \;=\; 0.$$

Therefore

$$\begin{aligned}
-\frac{1}{c}j^\nu(x)+\frac{1}{4\pi}\partial_\mu\left(\partial^\mu A^\nu\right)-\frac{1}{4\pi}\partial_\mu\left(\partial^\nu A^\mu\right) &= 0,\\
\frac{1}{4\pi}\partial_\mu F^{\mu\nu} &= \frac{1}{c}j^\nu(x),\\
\partial_\mu F^{\mu\nu} &= \frac{4\pi}{c}j^\nu(x).
\end{aligned}$$

4 Conservation laws

4.1 Noetherian conserved quantities

Consider the functional

$$I\left[\psi\right] = \int\limits_{\Sigma} d^4x\,\mathscr{L}\left(\psi(x),\partial_\mu\psi(x)\right).$$

Let us first consider a transformation of the fields, which leaves $\mathscr{L}$ strictly invariant. Assume that this transformation is given by

$$\psi(x) \quad\rightarrow\quad \psi'(x) = h^\alpha(\psi(x)),$$

with

$$h^0(\psi(x)) = \psi(x).$$

For α close to zero we have

$$\delta\psi = \psi' - \psi = \alpha\,\frac{d}{d\alpha}h^\alpha(\psi(x))\Big|_{\alpha=0}.$$

For the variation of the Lagrange density we obtain

$$\begin{aligned}
\delta\mathscr{L} &= \frac{\partial\mathscr{L}}{\partial\psi}\delta\psi + \frac{\partial\mathscr{L}}{\partial\left(\partial_\mu\psi\right)}\partial_\mu\delta\psi \\
&= \frac{\partial\mathscr{L}}{\partial\psi}\delta\psi + \partial_\mu\left(\frac{\partial\mathscr{L}}{\partial\left(\partial_\mu\psi\right)}\delta\psi\right) - \partial_\mu\left(\frac{\partial\mathscr{L}}{\partial\left(\partial_\mu\psi\right)}\right)\delta\psi \\
&= \left[\frac{\partial\mathscr{L}}{\partial\psi} - \partial_\mu\frac{\partial\mathscr{L}}{\partial\left(\partial_\mu\psi\right)}\right]\delta\psi + \partial_\mu\left(\frac{\partial\mathscr{L}}{\partial\left(\partial_\mu\psi\right)}\delta\psi\right)
\end{aligned}$$

If ψ is a solution of the Euler-Lagrange equations then the first term vanishes. Under the assumption that the Lagrange density is invariant under the transformation h^α, i.e. $\delta\mathscr{L} = 0$, it follows that then also the second term vanishes, e.g.

$$\partial_\mu J^\mu(x) = 0,$$

where the conserved current is given by

$$J^\mu(x) = \frac{\partial\mathscr{L}}{\partial\left(\partial_\mu\psi\right)}\delta\psi.$$

We may generalise Noether's theorem to transformations, which leave the Lagrange density invariant up to gauge terms, i.e. situations where we have

$$\mathscr{L}\left(A'_\mu,\partial_\mu A'_\nu\right) = \mathscr{L}\left(A_\mu,\partial_\mu A_\nu\right) + \frac{1}{c}j^\mu(x)\partial_\mu\Lambda(x)$$

instead of

$$\mathscr{L}\left(A'_\mu, \partial_\mu A'_\nu\right) \;=\; \mathscr{L}\left(A_\mu, \partial_\mu A_\nu\right).$$

For $\partial_\mu j^\mu = 0$ we may replace $j^\mu(x)\partial_\mu \Lambda(x)$ by

$$\partial_\mu \left(j^\mu(x)\Lambda(x)\right).$$

The additional term is a divergence and gives a surface term in the action integral. Since the variation of the fields vanishes there, the surface term yields zero and nothing changes.

4.2 Translational invariance and the energy-momentum tensor

Let us consider again the Lagrange density

$$\mathscr{L}\left(\psi(x), \partial_\mu \psi(x)\right),$$

which does not depend explicitly on x. Under translations

$$x^{\mu\prime} \;\to\; x^\mu + \alpha c^\mu,$$

we have

$$\psi(x) \;\to\; \psi'(x') = \psi(x+\alpha c) = \psi(x) + \delta\psi(x),$$

$$\delta\psi \;=\; \psi' - \psi = \alpha \left.\frac{d}{d\alpha}\psi(x+\alpha c)\right|_{\alpha=0} = \alpha c_\mu \partial^\mu \psi(x).$$

Furthermore

$$\delta\mathscr{L} \;=\; \mathscr{L}\left(\psi'(x'), \partial_\mu \psi'(x')\right) - \mathscr{L}\left(\psi(x), \partial_\mu \psi(x)\right) = \alpha c_\mu \partial^\mu \mathscr{L}\left(\psi(x), \partial_\nu \psi(x)\right).$$

Therefore we have

$$\delta\mathscr{L} \;=\; \partial_\mu \left(\frac{\partial \mathscr{L}}{\partial\left(\partial_\mu \psi\right)}\delta\psi\right) = \partial_\mu \left(\frac{\partial \mathscr{L}}{\partial\left(\partial_\mu \psi\right)}\alpha c_\mu \partial^\mu \psi(x)\right).$$

This implies

$$\begin{aligned}
\alpha c_\nu \partial^\nu \mathscr{L} - \partial_\mu \left(\frac{\partial \mathscr{L}}{\partial\left(\partial_\mu \psi\right)}\alpha c_\nu \partial^\nu \psi(x)\right) &= 0, \\
\alpha c_\nu \left[g^{\nu\mu}\partial_\mu \mathscr{L} - \partial_\mu \left(\frac{\partial \mathscr{L}}{\partial\left(\partial_\mu \psi\right)}\partial^\nu \psi(x)\right)\right] &= 0, \\
\alpha c_\nu \partial_\mu \left[g^{\nu\mu}\mathscr{L} - \left(\frac{\partial \mathscr{L}}{\partial\left(\partial_\mu \psi\right)}\partial^\nu \psi(x)\right)\right] &= 0,
\end{aligned}$$

We call the tensor field

$$T^{\mu\nu} = \left(\frac{\partial \mathscr{L}}{\partial \left(\partial_\mu \psi\right)} \partial^\nu \psi(x) \right) - g^{\mu\nu} \mathscr{L}$$

the **canonical energy-momentum tensor**. $T^{\mu\nu}$ satisfies the four conservation laws

$$\partial_\mu T^{\mu\nu} = 0.$$

Remark: If several fields $\psi^{(i)}$ appear in the Lagrange density, we sum over all fields:

$$T^{\mu\nu} = \sum_{i=1}^{N} \left(\frac{\partial \mathscr{L}}{\partial \left(\partial_\mu \psi^{(i)}\right)} \partial^\nu \psi^{(i)}(x) \right) - g^{\mu\nu} \mathscr{L}$$

Remark: If we add to $T^{\mu\nu}$ a term

$$\partial_\rho B^{\mu\rho\nu},$$

where $B^{\mu\rho\nu}$ is anti-symmetric in μ and ρ,

$$B^{\rho\mu\nu} = -B^{\mu\rho\nu},$$

we equally have

$$\partial_\mu \left(T^{\mu\nu} + \partial_\rho B^{\mu\rho\nu} \right) = 0.$$

This implies that the canonical energy-momentum tensor is not yet a unique conserved quantity. In order to arrive at a unique conserved quantity, one may consider in addition the angular momentum.

Preliminary remark: The relativistic generalisation of the angular momentum

$$\vec{M} = \vec{x} \times \vec{p}$$

is given by

$$M^{\mu\nu} = \frac{1}{2} \left(x^\mu p^\nu - x^\nu p^\mu \right).$$

We may impose on $T^{\mu\nu}$ the additional requirement that with the definition of the angular momentum density

$$M^{\mu\nu\rho} = T^{\mu\nu} x^\rho - T^{\mu\rho} x^\nu$$

we have

$$\partial_\mu M^{\mu\nu\rho} = 0.$$

This implies

$$\begin{aligned} \partial_\mu M^{\mu\nu\rho} &= \partial_\mu \left(T^{\mu\nu} x^\rho - T^{\mu\rho} x^\nu \right) = \left(\partial_\mu T^{\mu\nu} \right) x^\rho + T^{\rho\nu} - \left(\partial_\mu T^{\mu\rho} \right) x^\nu - T^{\nu\rho} \\ &= T^{\rho\nu} - T^{\nu\rho} = 0. \end{aligned}$$

Therefore

$$T^{\mu\nu} = T^{\nu\mu},$$

i.e. the energy-momentum tensor must be symmetric.

4.3 The energy-momentum tensor of the electromagnetic field

We consider the Lagrange density of the electromagnetic field without external sources:

$$\mathscr{L}\left(A_\mu, \partial_\mu A_\nu\right) = -\frac{1}{16\pi} F_{\mu\nu} F^{\mu\nu}$$

We obtain

$$\begin{aligned}\left(\frac{\partial \mathscr{L}}{\partial\left(\partial_\mu A_\tau\right)} \partial^\nu A_\tau\right) - g^{\mu\nu}\mathscr{L} &= -\frac{1}{4\pi}\left(\partial^\mu A^\tau\right)\left(\partial^\nu A_\tau\right) + \frac{1}{4\pi}\left(\partial^\tau A^\nu\right)\left(\partial^\nu A_\tau\right) + \frac{1}{16\pi} g^{\mu\nu} F_{\rho\sigma} F^{\rho\sigma} \\ &= \frac{1}{4\pi}\left[F^{\mu\tau}(x) F_\tau{}^\nu(x) + \frac{1}{4} g^{\mu\nu} F_{\rho\sigma} F^{\rho\sigma}\right] - \frac{1}{4\pi} F^{\mu\tau} \partial_\tau A^\nu.\end{aligned}$$

We are considering the case without external sources. This implies

$$\partial_\mu F^{\mu\nu} = 0,$$

and therefore

$$-\frac{1}{4\pi} F^{\mu\tau} \partial_\tau A^\nu = -\frac{1}{4\pi} \partial_\tau \left(F^{\mu\tau} A^\nu\right).$$

This term is a surface term. Therefore we find that the symmetric energy-momentum tensor of the electromagnetic field is given by

$$T^{\mu\nu} = \frac{1}{4\pi}\left[F^{\mu\tau}(x) F_\tau{}^\nu(x) + \frac{1}{4} g^{\mu\nu} F_{\rho\sigma} F^{\rho\sigma}\right].$$

Explicitly, we find for the individual components

$$\begin{aligned} T^{00} &= \frac{1}{8\pi}\left(\vec{E}^2 + \vec{B}^2\right) = u(t,\vec{x}), \\ T^{i0} &= \frac{1}{4\pi}\left(\vec{E} \times \vec{B}\right)^i = \frac{1}{c} S^i(t,\vec{x}), \\ T^{ij} &= -\frac{1}{4\pi}\left[\vec{E}^i \vec{E}^j + \vec{B}^i \vec{B}^j - \frac{1}{2}\delta^{ij}\left(\vec{E}^2 + \vec{B}^2\right)\right]. \end{aligned}$$

$u(t,\vec{x})$ denotes the **energy density** of the electromagnetic field. The vector $\vec{S}$ is called the **Poynting vector** and describes the momentum density (or the energy flux density). The purely spatial components T^{ij} are known as **Maxwell's stress tensor**.

Summary on Noether's theorem

$$\delta\mathscr{L} = \left[\frac{\partial\mathscr{L}}{\partial\psi} - \partial_\mu \frac{\partial\mathscr{L}}{\partial(\partial_\mu\psi)}\right]\delta\psi + \partial_\mu\left(\frac{\partial\mathscr{L}}{\partial(\partial_\mu\psi)}\delta\psi\right)$$

Case 1:
- ψ satisfies the Euler-Lagrange equations.
- $\mathscr{L}$ is strictly invariant under symmetry transformations.
Then: The Noether current

$$J^\mu(x) = \frac{\partial\mathscr{L}}{\partial(\partial_\mu\psi)}\delta\psi$$

is conserved:

$$\partial_\mu J^\mu(x) = 0.$$

Case 2:
- ψ satisfies the Euler-Lagrange equations.
- $\mathscr{L}$ is invariant under symmetry transformations up to gauge terms.
Then: The Noether current is also conserved.

Case 3:
- ψ satisfies the Euler-Lagrange equations.
- $\mathscr{L}$ does not depend explicitly on x_μ.
Then: The canonical energy-momentum tensor

$$T^{\mu\nu} = \left(\frac{\partial\mathscr{L}}{\partial(\partial_\mu\psi)}\partial^\nu\psi(x)\right) - g^{\mu\nu}\mathscr{L}$$

is conserved:

$$\partial_\mu T^{\mu\nu} = 0.$$

$T^{\mu\nu}$ is unique up to

$$T^{\mu\nu} \to T^{\mu\nu} + \partial_\rho B^{\mu\rho\nu}, \quad B^{\rho\mu\nu} = -B^{\mu\rho\nu},$$

Additional requirement: $T^{\mu\nu}$ is symmetric:

$$T^{\mu\nu} = T^{\nu\mu}.$$

Energy-momentum tensor of the electromagnetic field:

$$T^{\mu\nu} = \frac{1}{4\pi}\left[F^{\mu\tau}(x)F_\tau{}^\nu(x) + \frac{1}{4}g^{\mu\nu}F_{\rho\sigma}F^{\rho\sigma}\right].$$

5 Riemannian and semi-Riemannian geometry

5.1 Manifolds

A **topological space** is a set M together with a family $\mathcal{T}$ of subsets of M satisfying the following properties:

1. $\emptyset \in \mathcal{T}, M \in \mathcal{T}$
2. $U_1, U_2 \in \mathcal{T} \Rightarrow U_1 \cap U_2 \in \mathcal{T}$
3. For any index set A we have $U_\alpha \in \mathcal{T}; \alpha \in A \Rightarrow \bigcup\limits_{\alpha \in A} U_\alpha \in \mathcal{T}$

The sets $U \in \mathcal{T}$ are called **open**.

A topological space is called **Hausdorff** if for any two distinct points $p_1, p_2 \in M$ there exists open sets $U_1, U_2 \in \mathcal{T}$ with

$$p_1 \in U_1, \quad p_2 \in U_2, \quad U_1 \cap U_2 = \emptyset.$$

A map between topological spaces is called **continuous** if the pre-image of any open set is again open.

A bijective map which is continuous in both directions is called a **homeomorphism**.

An **open chart** on M is a pair (U, φ), where U is an open subset of M and φ is a homeomorphism of U onto an open subset of $\mathbb{R}^n$.

A **differentiable manifold** of dimension n is a Hausdorff space with a collection of open charts $(U_\alpha, \varphi_\alpha)_{\alpha \in A}$ such that

M1:

$$M = \bigcup_{\alpha \in A} U_\alpha.$$

M2: For each pair $\alpha, \beta \in A$ the mapping $\varphi_\beta \circ \varphi_\alpha^{-1}$ is an infinitely differentiable mapping of $\varphi_\alpha \left(U_\alpha \cap U_\beta\right)$ onto $\varphi_\beta \left(U_\alpha \cap U_\beta\right)$.

A differentiable manifold is also often denoted as a C^∞ manifold. As we will only be concerned with differentiable manifolds, we will often omit the word “differentiable” and just speak about manifolds.

The collection of open charts $(U_\alpha, \varphi_\alpha)_{\alpha \in A}$ is called an **atlas**.

If $p \in U_\alpha$ and

$$\varphi_\alpha(p) \;=\; (x_1(p), \ldots, x_n(p)),$$

the set U_α is called the **coordinate neighbourhood** of p and the numbers $x_i(p)$ are called the **local coordinates** of p.

Note that in each coordinate neighbourhood M looks like an open subset of $\mathbb{R}^n$. But note that we do not require that M be $\mathbb{R}^n$ globally.

Consider two manifolds M and N with dimensions m and n. Let x_i be coordinates on M and y_j be coordinates on N. A mapping $f : M \to N$ between two manifolds is called **analytic**, if for each point $p \in M$ there exits a neighbourhood U of p and n power series P_j, $j = 1, \ldots, n$ such that

$$y_j(f(q)) \;=\; P_j\left(x_1(q) - x_1(p), \ldots, x_m(q) - x_m(p)\right)$$

for all $q \in U$.

An **analytic manifold** is a manifold where the mapping $\varphi_\beta \circ \varphi_\alpha^{-1}$ is analytic.

Examples

a) $\mathbb{R}^n$: The space $\mathbb{R}^n$ is a manifold. $\mathbb{R}^n$ can be covered with a single chart.

b) S^1: The circle

$$S^1 = \{\vec{x} \in \mathbb{R}^2 | |\vec{x}|^2 = 1\}$$

is a manifold. For an atlas we need at least two charts.

c) S^n: The n-sphere, defined by

$$S^n = \{\vec{x} \in \mathbb{R}^{n+1} | |\vec{x}|^2 = 1\}$$

d) $\mathbb{P}^n(\mathbb{R})$: The projective space defined as all lines through the origin in $\mathbb{R}^{n+1}$:

$$(x_0, x_1, \ldots, x_n) \;=\; \lambda(x_0', x_1', \ldots, x_n'), \quad \lambda \neq 0.$$

e) The set of rotation matrices in two dimensions:

$$\left(\begin{array}{cc} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{array} \right),$$

The set of all these matrices forms a manifold homeomorphic to the circle S^1.

f) More generally, all Lie groups are by definition analytic manifolds.

Counterexamples

In order to understand better the definition of a manifold, let us give a few examples, which are **not** manifolds:

a) The union of a one-dimensional line with a two-dimensional surface. An example is given by

$$x_3 \left(x_1^2 + x_2^2\right) = 0.$$

This set is in a neighbourhood of some points homeomorph to $\mathbb{R}$, in the neighbourhood of other points homeomorph to $\mathbb{R}^2$. But the definition of a manifold requires that the set is at all points homeomorph to $\mathbb{R}^n$ for a fixed n.

b) The cone

$$x_1^2 + x_2^2 - x_3^2 = 0.$$

The neighbourhood of the point $(0,0,0)$ cannot be mapped homeomorphically to $\mathbb{R}^2$.

c) An individual cone segment

$$x_1^2 + x_2^2 - x_3^2 = 0, \qquad x_3 \geq 0.$$

Although we may map a neighbourhood of the point $(0,0,0)$ continuously to $\mathbb{R}^2$, this cannot be done in a differentiable way.

d) The line segment

$$[0,1].$$

The endpoints have no open neighbourhoods.

Morphisms

Let us summarise the various morphisms we encountered up to now:

Homeomorphism: A map $f : M \to N$ between two manifolds M and N is called a homeomorphism if it is bijective and both the mapping $f : M \to N$ and the inverse $f^{-1} : N \to M$ are continuous.

Diffeomorphism: A map $f : M \to N$ is called a diffeomorphism if it is a homeomorphism and both f and f^{-1} are infinitely differentiable.

Analytic diffeomorphism: The map $f : M \to N$ is a diffeomorphism and analytic.

5.2 Differential forms and integration on manifolds

Preliminary remark: We would like to define integrals on manifolds. The definition should on the one hand generalise volume integrals like

$$\int_M d^4x \, \mathscr{L}(x)$$

on an Euclidean space or on Minkowski space, and on the other hand also include line integrals as the one occurring for example in

$$-mc \int_a^b ds.$$

Let us first consider one-dimensional integrals, which we may define as the limit

$$\int_{\mathbb{R}} dx \, f(x) \;=\; \lim \sum_j f(x_j) \Delta x_j$$

In the same way we have for two-dimensional integrals:

$$\int_{\mathbb{R}^2} dx \, dy \, g(x,y) \;=\; \lim \sum_j \sum_k g(x_j, y_k) \Delta x_j \Delta y_k$$

Remark: The sign in the last example depends on the chosen orientation.

Instead of the functions $f(x)$ and $g(x,y)$ we will now introduce new objects

$$f(x) \, dx, \qquad g(x,y) \, dx \wedge dy,$$

which may be integrated over a domain of the appropriate dimension. The reason for introducing these new objects are the clearer transformation properties.

Tangent vectors

Let $I \subset \mathbb{R}$ be an interval and $\gamma : I \to M \subset \mathbb{R}^n$ a differentiable map. We call

$$\left. \frac{d}{dt} \gamma(t) \right|_{t_0} \in \mathbb{R}^n$$

a **tangent vector** on M at the point $\gamma(t_0)$. The set of all tangent vectors on M at the point p is called the **tangent space** T_pM at p. The dimension of the tangent space equals the dimension of the manifold.

We denote by T_p^*M the dual vector space of T_pM, i.e. the set of all linear maps

$$\phi : T_pM \to \mathbb{R}.$$

Elements $\phi \in T_p^*M$ are called **cotangent vectors** and T_p^*M is called the **cotangent space**. Linear maps from a vector space to $\mathbb{R}$ are also called linear forms.

A **vector field** is a map

$$X : M \to \bigcup_p T_p M$$

and associates to each point $p \in M$ a tangent vector $X(p) \in T_p M$.

Differential one-forms

A differential one-form is a map

$$\omega : M \to \bigcup_p T_p^* M$$

with $\omega(p) \in T_p^*M$. The differential one-form ω assigns to each point $p \in M$ a cotangent vector $\omega(p) \in T_p^*M$. We denote the value of $\omega(p)$ applied to the tangent vector $v \in T_p M$ by

$$\langle \omega(p), v \rangle.$$

Definition: Let $U \subset \mathbb{R}^n$ and let $f : U \to \mathbb{R}$ be a differentiable function. The total differential df of f is the differential one-form, which satisfies

$$\langle df(p), v \rangle \;=\; \sum_{i=1}^{n} \frac{\partial f(p)}{\partial x_i} v_i$$

for all tangent vectors $v = v_i e_i$.

With the help of the coordinate functions

$$x_i : \mathbb{R}^n \to \mathbb{R}, (y_1, ..., y_n) \to y_i$$

we may define the differentials

$$dx_1, ..., dx_n.$$

We have

$$\left\langle dx_i, e_j \right\rangle \;=\; \delta_{ij}.$$

The cotangent vectors $dx_1(p)$, ..., $dx_n(p)$ form a basis of T_p^*M.

Coordinate representation: Every differential one-form may be written as

$$\omega \;=\; \sum_{i=1}^{n} f_i(x) dx_i.$$

Line integrals: Let $\gamma : [a,b] \to U$ be a curve. We define the integral of ω along the curve γ by

$$\int_\gamma \omega \;=\; \int_a^b \langle \omega(\gamma(t)), \gamma'(t) \rangle dt.$$

Differential k-forms

We have seen that differential one-forms may be integrated along curves. We now seek a generalisation, which allows integration over domains of higher dimensions. We start with the definition of the wedge product for linear maps: Let ω_1, ..., $\omega_K \in V^*$ be linear forms, i.e.

$$\omega_j : V \to \mathbb{R}.$$

We define the map

$$\omega_1 \wedge ... \wedge \omega_k : V^k \to \mathbb{R}$$

by

$$(\omega_1 \wedge ... \wedge \omega_k)(v_1, ..., v_k) \;=\; \det \begin{pmatrix} \langle \omega_1, v_1 \rangle & ... & \langle \omega_1, v_k \rangle \\ ... & ... & ... \\ \langle \omega_k, v_1 \rangle & ... & \langle \omega_k, v_k \rangle \end{pmatrix}$$

Properties of the wedge product:

- The wedge product is linear in each argument:

$$\omega_1 \wedge ... \wedge \left(a\omega_i' + b\omega_i''\right) \wedge ... \wedge \omega_k = \\ a\left(\omega_1 \wedge ... \wedge \omega_i' \wedge ... \wedge \omega_k\right) + b\left(\omega_1 \wedge ... \wedge \omega_i'' \wedge ... \wedge \omega_k\right)$$

- The wedge product is alternating:

$$\omega_{\sigma(1)} \wedge ... \wedge \omega_{\sigma(k)} \;=\; \mathrm{sign}(\sigma) \cdot \omega_1 \wedge ... \wedge \omega_k$$

We denote the set of all alternating multilinear k-forms on V with

$$\wedge^k V^*.$$

Definition: A **differential k-form** is a map

$$\omega : M \to \bigcup_p \wedge^k T_p^* M$$

with $\omega(p) \in \wedge^k T_p^* M$. This definition coincides for $k = 1$ with the previous definition of a differential one-form. A differential 0-form is a real-valued function.

Coordinate representation of differential k-forms:

$$\begin{aligned} \omega &= \frac{1}{k!} \sum_{i_1,...,i_k} f_{i_1...i_k} dx_{i_1} \wedge ... \wedge dx_{i_k} \\ &= \sum_{i_1 < ... < i_k} f_{i_1...i_k} dx_{i_1} \wedge ... \wedge dx_{i_k}. \end{aligned}$$

Differentiation of differential forms: Let

$$\omega = \sum_{i_1<\ldots<i_k} f_{i_1\ldots i_k} dx_{i_1} \wedge \ldots \wedge dx_{i_k}.$$

be a k-form. We denote by $d\omega$ the differential $(k+1)$-form

$$d\omega = \sum_{i_1<\ldots<i_k} df_{i_1\ldots i_k} \wedge dx_{i_1} \wedge \ldots \wedge dx_{i_k}.$$

Rules: Let ω and ω' be two differential k-forms and let f be a function. Then $f\omega$ and $\omega+\omega'$, defined by

$$\begin{aligned} (f\omega)(p) &= f(p)\omega(p), \\ \left(\omega+\omega'\right)(p) &= \omega(p)+\omega'(p) \end{aligned}$$

are again differential k-forms. Furthermore, let σ be a differential l-form. We define a differential $(k+l)$-form $\omega \wedge \sigma$ by

$$(\omega \wedge \sigma)(p) = \omega(p) \wedge \sigma(p).$$

Remark:

$$\omega \wedge \sigma = (-1)^{kl} \sigma \wedge \omega.$$

We further have:

$$\begin{aligned} d\left(a\omega+b\omega'\right) &= ad\omega+bd\omega', \\ d(\omega \wedge \sigma) &= (d\omega) \wedge \sigma + (-1)^k \omega \wedge (d\sigma), \\ d(d\omega) &= 0. \end{aligned}$$

Pull-back of differential forms: Let $U \subset \mathbb{R}^n$ and let

$$\omega = \frac{1}{k!} \sum f_{i_1\ldots i_k} dx_{i_1} \wedge \ldots \wedge dx_{i_k}.$$

be a k-form on U. Let $V \subset \mathbb{R}^m$ be an open subset and consider a continuous differentiable map

$$\varphi = (\varphi_1, \ldots, \varphi_n) : V \to U.$$

We may define a differential k-form $\varphi^*\omega$ on V by

$$\varphi^*\omega = \frac{1}{k!} \sum \left(f_{i_1\ldots i_k} \circ \varphi\right) d\varphi_{i_1} \wedge \ldots \wedge d\varphi_{i_k}.$$

Remark: Differential k-forms may be integrated over k-dimensional (sub)-manifolds. Let M be a manifold of dimension n, let K be a submanifold of dimension k, and let A be a compact subset of K, also of dimension k. Further assume that ω is a differential k-form on M and

$$\varphi : U \to \mathbb{R}^n$$

a local chart of M such that $A \in U$. Then we have

$$\varphi^{-1} : \mathbb{R}^n \to U$$

and we define

$$\int_A \omega \;=\; \int_{\varphi(A)} \left(\varphi^{-1}\right)^* \omega$$

We pull-back the differential form ω by φ^{-1} to an open subset of $\mathbb{R}^n$. This reduces integration on manifolds to integration on $\mathbb{R}^n$.

Example: Consider the differential 2-form

$$\omega \;=\; 3x_3 dx_2 \wedge dx_3 + (x_1^2 + x_2^2) dx_3 \wedge dx_1 + x_1 x_3 dx_1 \wedge dx_2$$

on $\mathbb{R}^3$. Consider further the two-dimensional sub-manifold

$$M \;=\; \{(x_1, x_2, x_3) \in \mathbb{R}^3 : x_3 = x_1 x_2\}$$

and let A be the following compact subset of M:

$$A \;=\; \{(x_1, x_2, x_3) \in M : 0 \le x_1 \le 1, 0 \le x_2 \le 1\}.$$

We would like to compute

$$\int_A \omega.$$

We choose a local chart of M:

$$\begin{aligned} \varphi^{-1} \;&:\; \mathbb{R}^2 \to M, \\ & \quad (y_1, y_2) \to (y_1, y_2, y_1 y_2). \end{aligned}$$

The individual coordinate maps are

$$\left(\varphi^{-1}\right)_1 = y_1, \quad \left(\varphi^{-1}\right)_2 = y_2, \quad \left(\varphi^{-1}\right)_3 = y_1 y_2,$$

and therefore

$$d\left(\varphi^{-1}\right)_1 = dy_1, \quad d\left(\varphi^{-1}\right)_2 = dy_2, \quad d\left(\varphi^{-1}\right)_3 = y_2 dy_1 + y_1 dy_2.$$

Thus

$$\begin{aligned} \int_A \omega &= \int_{\varphi(A)} \left(\varphi^{-1}\right)^* \omega = \\ &= \int_{\varphi(A)} 3y_1 y_2 dy_2 \wedge (y_2 dy_1 + y_1 dy_2) + \left(y_1^2 + y_2^2\right)(y_2 dy_1 + y_1 dy_2) \wedge dy_1 + y_1 (y_1 y_2) dy_1 \wedge dy_2 \\ &= \int_{\varphi(A)} \left(y_1^2 y_2 - 4y_1 y_2^2 - y_1^3\right) dy_1 \wedge dy_2 = \int_0^1 dy_1 \int_0^1 dy_2 \left(y_1^2 y_2 - 4y_1 y_2^2 - y_1^3\right) = -\frac{3}{4}. \end{aligned}$$

We conclude the section on differential forms with examples occurring in physics: The gauge potential of electrodynamics defines a differential one-form

$$A = i\frac{e}{\hbar c}A_\mu(x)dx^\mu.$$

We further have

$$\begin{aligned} dA &= d\left(i\frac{e}{\hbar c}A_\nu dx^\nu\right) = i\frac{e}{\hbar c}\partial_\mu A_\nu dx^\mu \wedge dx^\nu \\ &= i\frac{e}{\hbar c}\frac{1}{2}\left(\partial_\mu A_\nu - \partial_\nu A_\mu\right) dx^\mu \wedge dx^\nu. \end{aligned}$$

This motivates to define a differential 2-form, related to the field strength by

$$F = dA = i\frac{e}{\hbar c}\frac{1}{2}F_{\mu\nu}dx^\mu \wedge dx^\nu.$$

Remark on the prefactors: We consider the following differential operator:

$$D_A = d + A = d + i\frac{e}{\hbar c}A_\mu dx^\mu = -\frac{i}{\hbar}\left(i\hbar d - \frac{q}{c}A_\mu dx^\mu\right).$$

Within quantum mechanics the term $i\hbar\partial_\mu$ corresponds to the momentum operator p_μ. We see that the term in the bracket is the four-dimensional generalisation of

$$\left(\vec{p} - \frac{q}{c}\vec{A}\right).$$

Finally, let us consider $D_A \wedge D_A$ applied to an arbitrary differential form ω:

$$\begin{aligned} (D_A \circ D_A)\,\omega &= \left(d + i\frac{e}{\hbar c}A_\mu dx^\mu\right) \circ \left(d + i\frac{e}{\hbar c}A_\nu dx^\nu\right)\omega \\ &= d\left(i\frac{e}{\hbar c}A_\mu dx^\mu \wedge \omega\right) + i\frac{e}{\hbar c}A_\nu dx^\nu \wedge d\omega - \left(\frac{e}{\hbar c}\right)^2 A_\mu A_\nu dx^\mu \wedge dx^\nu \wedge \omega \\ &= (dA) \wedge \omega \end{aligned}$$

Therefore

$$\begin{aligned} D_A &= d + A, \\ D_A^2 &= dA + A \wedge A = dA = F. \end{aligned}$$

D_A is called covariant derivative, F is called curvature form.

5.3 Tensors

We already defined tensors within special relativity. Let K and K' be two coordinate systems, related by a Lorentz transformation:

$$x'^\mu = \Lambda^\mu{}_\nu x^\nu.$$

We called a quantity $T^{\mu_1 \dots \mu_r}$, which transforms as

$$T'^{\mu_1 \dots \mu_r} = \Lambda^{\mu_1}_{\nu_1} \dots \Lambda^{\mu_1}_{\nu_1} T^{\nu_1 \dots \nu_r}$$

a rank r tensor. The contravariant four-vector x^μ is a rank 1 tensor.

We now generalise this definition to coordinate systems, which are related by an arbitrary coordinate transformation, i.e. not necessarily a Lorentz transformation. We consider the transformation from a coordinate system with coordinates x^0, x^1, x^2, x^3 to another coordinate system with coordinates x'^0, x'^1, x'^2, x'^3:

$$x'^\mu = f^\mu(x^0, x^1, x^2, x^3).$$

Under a change of coordinates, the differentials of the coordinates transform as

$$dx'^\mu = \frac{\partial x'^\mu}{\partial x^\nu} dx^\nu.$$

As **contravariant four-vector** we denote any set of four quantities A^μ ($\mu \in \{0,1,2,3\}$), which transform as these differentials under a change of coordinates:

$$A'^\mu = \frac{\partial x'^\mu}{\partial x^\nu} A^\nu.$$

Our main focus here are four-dimensional manifolds. Of course, there is a straightforward generalisation to D-dimensional manifolds, simply take $\mu \in \{0,1,\dots,D-1\}$.

This definition is compatible with the previous definition within special relativity, if the coordinate transformation is a Lorentz transformation: Let

$$x'^\mu = f^\mu(x^0, x^1, x^2, x^3) = \Lambda^\mu{}_\nu x^\nu.$$

Then

$$\frac{\partial x'^\mu}{\partial x^\nu} = \frac{\partial f^\mu(x^0, x^1, x^2, x^3)}{\partial x^\nu} = \Lambda^\mu{}_\nu$$

and therefore

$$x'^\mu = \Lambda^\mu{}_\nu x^\nu = \frac{\partial x'^\mu}{\partial x^\nu} x^\nu.$$

Let ϕ be a scalar function. The derivatives $\partial\phi/\partial x^\mu$ transform under a change of coordinates as

$$\frac{\partial \phi}{\partial x'^\mu} = \frac{\partial \phi}{\partial x^\nu} \frac{\partial x^\nu}{\partial x'^\mu}.$$

We call any set of four quantities A_μ ($\mu \in \{0,1,2,3\}$), which transform under a change of coordinates as the derivatives of a scalar function a **covariant four-vector**:

$$A'_\mu = \frac{\partial x^\nu}{\partial x'^\mu} A_\nu$$

We may write a tangent vector at any point as a linear combination of basis vectors e_μ:

$$V = V^\mu e_\mu.$$

Sometimes, an alternative notation for the basis vectors of the tangent space is used:

$$\partial_\mu = e_\mu.$$

(It should be clear from the context if ∂_μ denotes a partial derivative or a basis vector of the tangent space.)

A vector field assigns to every point of a manifold a vector. The dual of a vector field is a one-form. A one-form assigns at every point of the manifold to a vector a (real or complex) number, or phrased differently, a one-form assigns to every point of the manifold a cotangent vector. A basis for the space of cotangent vectors is given by the differentials dx^μ:

$$\omega = \omega_\mu dx^\mu.$$

Duality between vector fields and one-forms implies

$$dx^\mu(\partial_\nu) = \delta^\mu_\nu.$$

Due to this duality we may re-interpret a vector field as follows: Originally, we defined a vector field as a map, which assigns to every point of the manifold a tangent vector. With the help of the duality we may equally well view a vector field as a map, which assigns to every point of the manifold a linear form, which in turn maps a cotangent vector to $\mathbb{R}$.

A tensor field with r contravariant and s covariant indices maps at the point $x \in M$ r r cotangent vectors and s tangent vectors to a real number.

$$\begin{aligned} (T^r_s)_x \; : \;& (T^*_x M)^r \times (T_x M)^s \to \mathbb{R}, \\ & \omega^1, ..., \omega^r, V_1, ..., V_s \to (T^r_s)_x \left(\omega^1, ..., \omega^r, V_1, ..., V_s\right). \end{aligned}$$

Coordinate representation:

$$t^{\mu_1,...,\mu_r}_{\nu_1,...,\nu_s}(x) = (T^r_s)_x \left(dx^{\mu_1}, ..., dx^{\mu_r}, \partial_{\nu_1}, ..., \partial_{\nu_s}\right).$$

Basis representation of a tensor field on a D-dimensional manifold (where the coordinates are indexed from 0 to $D-1$):

$$T^r_s = \sum_{\mu_1,...,\mu_r=0}^{D-1} \sum_{\nu_1,...,\nu_s=0}^{D-1} t^{\mu_1,...,\mu_r}_{\nu_1,...,\nu_s}(x) \left(\partial_{\mu_1} \otimes ... \otimes \partial_{\mu_r}\right) \otimes \left(dx^{\nu_1} \otimes ... \otimes dx^{\nu_s}\right).$$

Example: A $(0,2)$-tensor field is given by

$$g = \sum_{\mu,\nu=0}^{D-1} g_{\mu\nu}(x) dx^\mu \otimes dx^\nu.$$

Remark: For a general $(0,s)$-tensor field the tensor product $\otimes$ appears, not the wedge product $\wedge$. Differential forms have the additional property of being anti-symmetric and we have

$$dx^\mu \wedge dx^\nu = \frac{1}{2}\left(dx^\mu \otimes dx^\nu - dx^\nu \otimes dx^\mu\right),$$

and more generally

$$dx^{\mu_1} \wedge dx^{\mu_2} \wedge \cdots \wedge dx^{\mu_k} = \frac{1}{k!}\sum_{\sigma \in S_k} \mathrm{sign}(\sigma)\; dx^{\mu_{\sigma(1)}} \otimes dx^{\mu_{\sigma(2)}} \otimes \cdots \otimes dx^{\mu_{\sigma(k)}}.$$

5.4 Riemannian manifolds

Definition of a Riemannian manifold: Let M be a differentiable manifold. A **Riemannian metric** g on M is a $(0,2)$-tensor field on M, such that for every point $x \in M$ we have:

$$\begin{aligned} g_x(U,V) &= g_x(V,U) \\ g_x(U,U) &\geq 0, \text{ and } g_x(U,U) = 0 \text{ only for } U = 0, \end{aligned}$$

where $U,V \in T_xM$ and $g_x = g|_x$.
In short this means that g_x is a symmetric positive-definite bilinear form. A manifold with a Riemannian metric is called a **Riemannian manifold**.

A $(0,2)$-tensor field g on M is called **semi-Riemannian metric** if

$$\begin{aligned} g_x(U,V) &= g_x(V,U), \\ \text{if } g_x(U,V) &= 0 \text{ for all } U \in T_xM, \text{ then } V = 0. \end{aligned}$$

A manifold with a semi-Riemannian metric is called a **semi-Riemannian manifold**.

Let (U,φ) be a chart of M and let $\{x^\mu\}$ be local coordinates. The metric is written as

$$g_x = g_{\mu\nu}(x) dx^\mu \otimes dx^\nu,$$

where we used Einstein's summation convention.

Remark: Since the metric is symmetric, the eigenvalues of $g_{\mu\nu}$ are real. For a Riemannian metric all eigenvalues are positive. For a semi-Riemannian metric the eigenvalues are positive or negative (and non-zero). A Manifold, where $g_{\mu\nu}$ has exactly one positive eigenvalue (and hence $(D-1)$ negative eigenvalues) is called a **Lorentz manifold**.

Let us elaborate on the notation: Instead of $g_{\mu\nu}(x)dx^\mu \otimes dx^\nu$ the notation

$$g_{\mu\nu}(x) dx^\mu dx^\nu$$

is frequently used, where the symbol $\otimes$ for the tensor product has been dropped. Also in this shortened notation the differentials dx^μ denote a basis of the cotangent space and $g = g_{\mu\nu}(x) dx^\mu dx^\nu$ is a $(0,2)$-tensor field.

A further notation is

$$g = \det\left(g_{\mu\nu}\right),$$

and

$$|g| = \left|\det\left(g_{\mu\nu}\right)\right|.$$

It should be clear from the context, if g denotes the $(0,2)$-tensor field $g_{\mu\nu}(x) dx^\mu dx^\nu$ or the determinant $\det(g_{\mu\nu})$.

The inverse of $g_{\mu\nu}$ is denoted by $g^{\mu\nu}$:

$$g_{\mu\rho} g^{\rho\nu} = g^{\nu\rho} g_{\rho\mu} = \delta_\mu^\nu.$$

The metric induces an isomorphism between $T_x M$ and $T_x^* M$. This isomorphism is explicitly given by

$$\begin{aligned} T_x M &\to T_x^* M, \\ U^\mu \partial_\mu &\to \left(U^\mu g_{\mu\nu}\right) dx^\nu \end{aligned}$$

and

$$\begin{aligned} T_x^* M &\to T_x M, \\ \omega_\mu dx^\mu &\to \left(\omega_\mu g^{\mu\nu}\right) \partial_\nu. \end{aligned}$$

Let us further discuss **tensor densities**. We recall the definition of the total anti-symmetric tensor (i.e. the Levi-Civita tensor):

$$\begin{aligned} \varepsilon_{\mu_1\mu_2\ldots\mu_n} &= 1 \quad \text{if } \mu_1, \mu_2, \ldots, \mu_n \text{ is an even permutation of } 0, 1, \ldots, (n-1), \\ \varepsilon_{\mu_1\mu_2\ldots\mu_n} &= -1 \quad \text{if } \mu_1, \mu_2, \ldots, \mu_n \text{ is an odd permutation of } 0, 1, \ldots, (n-1), \\ \varepsilon_{\mu_1\mu_2\ldots\mu_n} &= 0 \quad \text{otherwise.} \end{aligned}$$

In flat Minkowski space the Levi-Civita symbol $\varepsilon_{\mu\nu\rho\sigma}$ transforms as a pseudotensor. Let us study, how the Levi-Civita symbol transforms on arbitrary manifolds. Let $M^\mu{}_{\mu'}$ be an arbitrary $n \times n$-matrix and denote $|M| = \det M^\mu{}_{\mu'}$. We have

$$\varepsilon_{\mu_1'\mu_2'\ldots\mu_n'} |M| = \varepsilon_{\mu_1\mu_2\ldots\mu_n} M^{\mu_1}{}_{\mu_1'} M^{\mu_2}{}_{\mu_2'} \ldots M^{\mu_n}{}_{\mu_n'}.$$

If we now take

$$M^\mu{}_{\mu'} = \frac{\partial x^\mu}{\partial x^{\mu'}}$$

we obtain

$$\varepsilon_{\mu_1'\mu_2'...\mu_n'} = \left|\frac{\partial x^{\mu'}}{\partial x^{\mu}}\right| \varepsilon_{\mu_1\mu_2...\mu_n} \frac{\partial x^{\mu_1}}{\partial x^{\mu_1'}} \frac{\partial x^{\mu_2}}{\partial x^{\mu_2'}} \cdots \frac{\partial x^{\mu_n}}{\partial x^{\mu_n'}}.$$

This is almost the transformation law of a rank n tensor. The transformation law is spoiled by the appearance of the determinant $|\partial x^{\mu'}/\partial x^{\mu}|$.
Let us further consider the transformation law of $g = \det g_{\mu\nu}$. One finds

$$g(x') = \left|\frac{\partial x^{\mu'}}{\partial x^{\mu}}\right|^{-2} g(x).$$

In general, we call a quantity, which transforms as

$$\left|\frac{\partial x^{\mu'}}{\partial x^{\mu}}\right|^{m} \times \text{Tensor}$$

a **tensor density of weight** m. We see that $\varepsilon_{\mu_1\mu_2...\mu_n}$ is a tensor density of weight 1 and g is a tensor density of weight (-2). The combination

$$\sqrt{|g|}\, \varepsilon_{\mu_1\mu_2...\mu_n}$$

transforms as a tensor.

Let us conclude this section by giving a useful formula for the contraction of two Levi-Civita symbols. We have

$$|g|\, \varepsilon_{\mu_1\mu_2...\mu_r\sigma_1...\sigma_{n-r}} \varepsilon^{\nu_1\nu_2...\nu_r\sigma_1...\sigma_{n-r}} = (-1)^s (n-r)! \delta^{\nu_1\nu_2...\nu_r}_{\mu_1\mu_2...\mu_r},$$

where s denotes the number of negative eigenvalues of the metric and

$$\delta^{\nu_1\nu_2...\nu_r}_{\mu_1\mu_2...\mu_r} = \begin{vmatrix} \delta^{\nu_1}_{\mu_1} & ... & \delta^{\nu_r}_{\mu_1} \\ ... & ... & ... \\ \delta^{\nu_1}_{\mu_r} & ... & \delta^{\nu_r}_{\mu_r} \end{vmatrix}.$$

5.5 Hodge theory

5.5.1 The Hodge $*$-operator

Let M be a m-dimensional manifold. If M is equipped with a metric, there is a natural isomorphism between the space of all differential r forms and the space of all differential $(m-r)$ forms, given by the Hodge $*$-operator (pronounce "Hodge star operator"):

$$\begin{aligned} * \ : \ & \Omega^r(M) \to \Omega^{m-r}(M) \\ & *(dx^{\mu_1} \wedge ... \wedge dx^{\mu_r}) = \frac{\sqrt{|g|}}{(m-r)!} \varepsilon^{\mu_1...\mu_r}{}_{\nu_{r+1}...\nu_m} dx^{\nu_{r+1}} \wedge ... \wedge dx^{\nu_m} \end{aligned}$$

Remark:

$$* * \omega = (-1)^{r(m-r)+s} \omega,$$

where s denotes the number of negative eigenvalues of the metric. This formula is easily verified by considering

$$\begin{aligned}
* * (dx^{\mu_1} \wedge ... \wedge dx^{\mu_r}) &= \frac{|g|}{r!\,(m-r)!} \varepsilon^{\mu_1 ... \mu_r}{}_{\sigma_{r+1} ... \sigma_m} \varepsilon^{\sigma_{r+1} ... \sigma_m}{}_{\nu_1 ... \nu_r} (dx^{\nu_1} \wedge ... \wedge dx^{\nu_r}) \\
&= (-1)^{r(m-r)} \frac{|g|}{r!\,(m-r)!} \varepsilon^{\mu_1 ... \mu_r \sigma_{r+1} ... \sigma_m} \varepsilon_{\nu_1 ... \nu_r \sigma_{r+1} ... \sigma_m} (dx^{\nu_1} \wedge ... \wedge dx^{\nu_r}) \\
&= \frac{(-1)^{r(m-r)+s}}{r!} \delta^{\mu_1 ... \mu_r}_{\nu_1 ... \nu_r} (dx^{\nu_1} \wedge ... \wedge dx^{\nu_r}) \\
&= (-1)^{r(m-r)+s} (dx^{\mu_1} \wedge ... \wedge dx^{\mu_r}).
\end{aligned}$$

The Hodge $*$-operator allows to define a scalar product between two r forms. Let

$$\begin{aligned}
\omega &= \frac{1}{r!} \omega_{\mu_1 ... \mu_r} dx^{\mu_1} \wedge ... \wedge dx^{\mu_k}, \\
\eta &= \frac{1}{r!} \eta_{\mu_1 ... \mu_r} dx^{\mu_1} \wedge ... \wedge dx^{\mu_k}.
\end{aligned}$$

One sets

$$\begin{aligned}
(\omega, \eta) &= \int_M \omega \wedge * \eta \\
&= \frac{1}{r!} \int_M \omega_{\mu_1 ... \mu_r} \eta^{\mu_1 ... \mu_r} \sqrt{|g|} dx^1 \wedge ... \wedge dx^m.
\end{aligned}$$

This product is symmetric:

$$(\omega, \eta) = (\eta, \omega)$$

Example:

$$*F = * \left(i \frac{e}{\hbar c} \frac{1}{2} F_{\mu\nu} dx^\mu \wedge dx^\nu \right) = \frac{1}{4} i \frac{e}{\hbar c} F^{\mu\nu} \varepsilon_{\mu\nu\rho\sigma} dx^\rho \wedge dx^\sigma = \left(i \frac{e}{\hbar c} \right) \frac{1}{2} \tilde{F}_{\mu\nu} dx^\mu \wedge dx^\nu.$$

We further have

$$(F, F) = \frac{1}{2} \left(i \frac{e}{\hbar c} \right)^2 \int d^4 x F_{\mu\nu} F^{\mu\nu}$$

and therefore

$$\int d^4 x \, \mathscr{L} = \frac{1}{8\pi} \left(\frac{\hbar c}{e} \right)^2 (F, F).$$

5.5.2 Self dual and anti-self dual forms

Let us consider the special case, where the manifold M is of even dimension $m = 2r$. In this case, the Hodge $*$-operator maps a r form to a r form.

Of particular interest is the case $m = 4$ and $r = 2$. Let

$$\omega = \frac{1}{2}\omega_{\mu\nu} dx^\mu \wedge dx^\nu$$

be a two-form. On a four-dimensional Lorentz manifold we have

$$**\omega = -\omega.$$

Let us now consider complex-valued differential forms. We call a two-form on a four-dimensional Lorentz manifold **self dual** if

$$i*\omega = \omega,$$

and **anti-self dual** if

$$i*\omega = -\omega.$$

The factor i is required to satisfy in both cases $**\omega = -\omega$. In the case of a four-dimensional Euclidean manifold the factor i does not appear.

In terms of components we have

$$*\omega = \frac{1}{2}\tilde{\omega}_{\mu\nu} dx^\mu \wedge dx^\nu, \qquad \tilde{\omega}_{\mu\nu} = \frac{1}{2}\sqrt{|g|}\omega_{\rho\sigma}\varepsilon^{\rho\sigma}{}_{\mu\nu}.$$

The conditions for being self dual or anti-self dual translate to

$$\begin{aligned} \text{self dual} \;:\; \omega_{\mu\nu} &= \frac{i}{2}\sqrt{|g|}\omega_{\rho\sigma}\varepsilon^{\rho\sigma}{}_{\mu\nu}, \\ \text{anti-self dual} \;:\; \omega_{\mu\nu} &= -\frac{i}{2}\sqrt{|g|}\omega_{\rho\sigma}\varepsilon^{\rho\sigma}{}_{\mu\nu}. \end{aligned}$$

An arbitrary two-form can always be decomposed into a self dual part and an anti-self dual part:

$$\omega = \omega^{\text{selfdual}} + \omega^{\text{antiselfdual}},$$

with

$$\begin{aligned} \omega^{\text{selfdual}} &= \frac{1}{2}\left(\omega + i*\omega\right), \\ \omega^{\text{antiselfdual}} &= \frac{1}{2}\left(\omega - i*\omega\right). \end{aligned}$$

With

$$\omega^{\text{selfdual}} = \frac{1}{2}\omega^{\text{selfdual}}_{\mu\nu} dx^\mu \wedge dx^\nu, \qquad \omega^{\text{antiselfdual}} = \frac{1}{2}\omega^{\text{antiselfdual}}_{\mu\nu} dx^\mu \wedge dx^\nu$$

we obtain

$$\begin{aligned} \omega^{\text{selfdual}}_{\mu\nu} &= \frac{1}{2}\left(\omega_{\mu\nu} + i\tilde{\omega}_{\mu\nu}\right) = \frac{1}{2}\left(\omega_{\mu\nu} + \frac{i}{2}\sqrt{|g|}\omega_{\rho\sigma}\varepsilon^{\rho\sigma}{}_{\mu\nu}\right), \\ \omega^{\text{antiselfdual}}_{\mu\nu} &= \frac{1}{2}\left(\omega_{\mu\nu} - i\tilde{\omega}_{\mu\nu}\right) = \frac{1}{2}\left(\omega_{\mu\nu} - \frac{i}{2}\sqrt{|g|}\omega_{\rho\sigma}\varepsilon^{\rho\sigma}{}_{\mu\nu}\right). \end{aligned}$$

5.6 The covariant derivative

In a flat space the derivatives of a vector

$$\frac{\partial}{\partial x^\nu} A_\mu$$

form a tensor. However, this is no longer true in a curved space, as one compares a vector at two different points.

Definition of an affine connection: An **affine connection** is a map ∇

$$\begin{aligned} \nabla \quad : \quad & \mathrm{Vect}(M) \times \mathrm{Vect}(M) \rightarrow \mathrm{Vect}(M) \\ & (X,Y) \rightarrow \nabla_X Y, \end{aligned}$$

which satisfies

$$\begin{aligned} \nabla_{(X+Y)} Z &= \nabla_X Z + \nabla_Y Z, \\ \nabla_{(fX)} Y &= f \nabla_X Y, \\ \nabla_X (Y+Z) &= \nabla_X Y + \nabla_X Z, \\ \nabla_X (fY) &= X(f) Y + f \nabla_X Y, \end{aligned}$$

where $f \in F(M)$ and $X, Y, Z \in \mathrm{Vect}(M)$.
Let (U, φ) be a chart with coordinates $x = \varphi(p)$. We define D^3 functions $C^\mu{}_{\nu\lambda}$ called **connection coefficients** by

$$\nabla_{e_\mu} e_\nu \;=\; e_\lambda C^\lambda{}_{\mu\nu},$$

where $\{e_\mu\} = \{\partial/\partial_\mu\}$ denotes the coordinate basis of T_pM. For functions $f \in F(M)$ we define

$$\nabla_X f \;=\; X(f) = X^\mu \left(\frac{\partial f}{\partial x^\mu} \right).$$

Then $\nabla_X(fY)$ takes the form of the Leibniz rule

$$\nabla_X (fY) \;=\; (\nabla_X f) Y + f (\nabla_X Y).$$

We further set for tensors

$$\nabla_X (T_1 \otimes T_2) \;=\; (\nabla_X T_1) \otimes T_2 + T_1 \otimes (\nabla_X T_2).$$

In the following we will use the notation

$$\nabla_\mu \;=\; \nabla_{e_\mu}.$$

Note that

$$\begin{aligned}
\nabla_X Y &= X^\mu \nabla_\mu \left(Y^\nu e_\nu\right) = X^\mu \left(\frac{\partial Y^\nu}{\partial x^\mu} e_\nu + Y^\nu \nabla_\mu e_\nu\right) \\
&= X^\mu \left(\frac{\partial Y^\lambda}{\partial x^\mu} + Y^\nu C^\lambda_{\mu\nu}\right) e_\lambda .
\end{aligned}$$

$\nabla_X Y$ is independent of the derivative of X. This motivates to consider

$$\nabla_\mu = \nabla_{\hat{e}_\mu}.$$

∇_μ is called the **covariant derivative**. We may re-write the above equation as follows:

$$\nabla_\mu \left(Y^\nu e_\nu\right) = \left(\partial_\mu Y^\nu + C^\nu_{\mu\lambda} Y^\lambda\right) e_\nu$$

Within the physics literature the basis vector e_ν is often dropped and one encounters for the components the notation:

$$\nabla_\mu Y^\nu = \partial_\mu Y^\nu + C^\nu_{\mu\lambda} Y^\lambda$$

We should always interpret this equation as if the missing basis vector is present. In strict mathematical terms we have

$$\begin{aligned}
\nabla_\mu e_\nu &= C^\lambda_{\mu\nu} e_\lambda, \\
\nabla_\mu Y^\nu &= \partial_\mu Y^\nu, \\
\nabla_\mu \left(Y^\nu e_\nu\right) &= \left(\partial_\mu Y^\nu + C^\nu_{\mu\lambda} Y^\lambda\right) e_\nu .
\end{aligned}$$

Let us also consider the action of the covariant derivative on covariant indices. Let $\omega = \omega_\mu dx^\mu$ and $Y = Y^\nu e_\nu$. We have

$$\nabla_\mu \langle \omega, Y \rangle = \nabla_\mu \left(\omega_\nu Y^\nu\right) = \left(\partial_\mu \omega_\nu\right) Y^\nu + \omega_\nu \left(\partial_\mu Y^\nu\right).$$

On the other hand we must have

$$\begin{aligned}
\nabla_\mu \langle \omega, Y \rangle &= \left\langle \nabla_\mu \omega, Y \right\rangle + \left\langle \omega, \nabla_\mu Y \right\rangle \\
&= \left\langle (\partial_\mu \omega_\nu) dx^\nu + \omega_\nu \nabla_\mu dx^\nu, Y \right\rangle + \left\langle \omega, (\partial_\mu Y^\nu + C^\nu_{\mu\lambda} Y^\lambda) e_\nu \right\rangle \\
&= (\partial_\mu \omega_\nu) Y^\nu + \left\langle \omega_\nu \nabla_\mu dx^\nu, Y^\lambda e_\lambda \right\rangle + \omega_\nu (\partial_\mu Y^\nu + C^\nu_{\mu\lambda} Y^\lambda).
\end{aligned}$$

Therefore

$$\omega_\nu \left\langle \nabla_\mu dx^\nu, e_\lambda \right\rangle Y^\lambda + \omega_\nu C^\nu_{\mu\lambda} Y^\lambda = 0$$

and hence

$$\nabla_\mu dx^\nu = -C^\nu_{\mu\lambda} dx^\lambda .$$

Therefore we have

$$\nabla_\mu \left(\omega_\nu dx^\nu\right) = \left(\partial_\mu \omega_\nu - C^\lambda_{\mu\nu}\omega_\lambda\right) dx^\nu.$$

Also in this case one finds in the physics literature the notation

$$\nabla_\mu \omega_\nu = \partial_\mu \omega_\nu - C^\lambda_{\mu\nu}\omega_\lambda$$

As in the case above we have to interpret this equation as if the missing basis vector dx^ν is present.

Parallel transport: If

$$\nabla_V X = 0$$

we say that the vector X is parallel transported along the curve defined by V.

5.7 The Levi-Civita connection

If a manifold is equipped with a metric, we may impose additional requirements on the affine connection: The first condition that we will impose is that the metric $g_{\mu\nu}$ is covariantly constant, i.e. we require that if two vectors X and Y are parallel transported along a curve the scalar product between the two vectors does not change. We may express this by the formula

$$\nabla_V \left(g(X,Y)\right) = 0,$$

for all X and Y with $\nabla_V X = \nabla_V Y = 0$. Since this holds for all curves and all parallel transported vectors, it follows that

$$\nabla_\kappa \left(g_{\mu\nu} dx^\mu \otimes dx^\nu\right) = 0,$$

or equivalently

$$\left(\partial_\kappa g_{\mu\nu} - C^\lambda_{\kappa\mu} g_{\lambda\nu} - C^\lambda_{\kappa\nu} g_{\mu\lambda}\right) dx^\mu \otimes dx^\nu = 0.$$

This has to hold for all components and therefore it follows that

$$\partial_\kappa g_{\mu\nu} - C^\lambda_{\kappa\mu} g_{\lambda\nu} - C^\lambda_{\kappa\nu} g_{\mu\lambda} = 0.$$

This is also written as

$$\nabla_\kappa g_{\mu\nu} = 0.$$

In this case we may write the connection coefficients $C^\kappa{}_{\mu\nu}$ as

$$C^\kappa{}_{\mu\nu} = \Gamma^\kappa{}_{\mu\nu} + K^\kappa{}_{\mu\nu}.$$

The quantities $\Gamma^{\kappa}{}_{\mu\nu}$ are called **Christoffel symbols**, They are symmetric in $\mu \leftrightarrow \nu$. The quantities $K^{\kappa}_{\mu\nu}$ are called **contorsion coefficients**. The explicit expressions for these quantities are

$$\begin{aligned}
\Gamma^{\kappa}{}_{\mu\nu} &= \frac{1}{2} g^{\kappa\lambda} \left(\partial_\mu g_{\nu\lambda} + \partial_\nu g_{\mu\lambda} - \partial_\lambda g_{\mu\nu} \right), \\
K^{\kappa}_{\mu\nu} &= \frac{1}{2} \left(T^{\kappa}_{\mu\nu} + T_{\mu}{}^{\kappa}{}_{\nu} + T_{\nu}{}^{\kappa}{}_{\mu} \right), \\
T^{\kappa}_{\mu\nu} &= C^{\kappa}_{\mu\nu} - C^{\kappa}_{\nu\mu}.
\end{aligned}$$

$T^{\kappa}_{\mu\nu}$ is anti-symmetric in $\mu \leftrightarrow \nu$. It can be shown that the quantities $T^{\kappa}_{\mu\nu}$ define a tensor, which is called the **torsion tensor**.

An affine connection is called **symmetric**, if the torsion tensor vanishes. In this case we have

$$\begin{aligned}
C^{\kappa}_{\mu\nu} &= \Gamma^{\kappa}{}_{\mu\nu}, \\
\Gamma^{\lambda}_{\mu\nu} &= \Gamma^{\lambda}_{\nu\mu}.
\end{aligned}$$

This is the second condition which we will impose: We require that the affine connection is symmetric, i.e. that the torsion tensor vanishes.

Theorem: On a Riemannian manifold or semi-Riemannian manifold (M, g) there is a unique symmetric connection, which is compatible with the metric (i.e. the metric is covariantly constant). This connection is called the **Levi-Civita connection**.

Assuming that the metric is covariantly constant and assuming that the connection coefficients are symmetric, we may easily derive the formula for the Christoffel symbols, thus proving the existence and uniqueness. We start by writing down the equation which expresses that the metric is covariantly constants for three different permutations of indices:

$$\begin{aligned}
\nabla_\rho g_{\mu\nu} &= \partial_\rho g_{\mu\nu} - \Gamma^{\lambda}_{\rho\mu} g_{\lambda\nu} - \Gamma^{\lambda}_{\rho\nu} g_{\mu\lambda} = 0, \\
\nabla_\mu g_{\nu\rho} &= \partial_\mu g_{\nu\rho} - \Gamma^{\lambda}_{\mu\nu} g_{\lambda\rho} - \Gamma^{\lambda}_{\mu\rho} g_{\nu\lambda} = 0, \\
\nabla_\nu g_{\rho\mu} &= \partial_\nu g_{\rho\mu} - \Gamma^{\lambda}_{\nu\rho} g_{\lambda\mu} - \Gamma^{\lambda}_{\nu\mu} g_{\rho\lambda} = 0.
\end{aligned}$$

If we subtract the last two equations from the first one we obtain

$$\partial_\rho g_{\mu\nu} - \partial_\mu g_{\nu\rho} - \partial_\nu g_{\rho\mu} + \Gamma^{\lambda}_{\mu\nu} g_{\lambda\rho} + \Gamma^{\lambda}_{\nu\mu} g_{\rho\lambda} + \Gamma^{\lambda}_{\nu\rho} g_{\lambda\mu} - \Gamma^{\lambda}_{\rho\nu} g_{\mu\lambda} + \Gamma^{\lambda}_{\mu\rho} g_{\nu\lambda} - \Gamma^{\lambda}_{\rho\mu} g_{\lambda\nu} = 0.$$

We now use the symmetry of the metric and of the Christoffel symbols. We obtain

$$\partial_\rho g_{\mu\nu} - \partial_\mu g_{\nu\rho} - \partial_\nu g_{\rho\mu} + 2\Gamma^{\lambda}_{\mu\nu} g_{\lambda\rho} = 0.$$

Solving for the Christoffel symbol we obtain the formula

$$\Gamma^{\kappa}{}_{\mu\nu} = \frac{1}{2} g^{\kappa\lambda} \left(\partial_\mu g_{\nu\lambda} + \partial_\nu g_{\mu\lambda} - \partial_\lambda g_{\mu\nu} \right).$$

5.8 Stokes' theorem

Stokes' theorem may be written elegantly with the help of differential forms on a differentiable manifold with a boundary as

$$\int\limits_M d\omega = \int\limits_{\partial M} \omega.$$

Here, M denotes a n-dimensional manifold, which may have a boundary. The boundary is denoted by ∂M and ω denotes a differential $(n-1)$-form.

If the manifold is endowed with a metric g, we may re-write Stokes' theorem as follows:

$$\int\limits_M d^n x \sqrt{|g|} \nabla_\mu V^\mu = \int\limits_{\partial M} d^{n-1} y \sqrt{|\gamma|}\, n_\mu V^\mu.$$

Here we denote by ∇_μ the covariant derivative with respect to the Levi-Civita connection, we denote by γ the metric on ∂M induced by g and we denote by n_μ a unit normal vector on ∂M.

The second version of Stokes' theorem is derived from the first version of Stokes' theorem for semi-Riemannian manifolds as follows: Since M is equipped with a metric, we may write any differential $(n-1)$-form as the Hodge dual of a differential one-form $V = V_\mu dx^\mu$:

$$\omega = *V$$

With

$$\omega = \frac{1}{(n-1)!} \omega_{\mu_1 \dots \mu_{n-1}} dx^{\mu_1} \wedge \dots \wedge dx^{\mu_{n-1}}$$

we have

$$\omega_{\mu_1 \dots \mu_{n-1}} = \sqrt{|g|} V_\mu g^{\mu\nu} \varepsilon_{\nu\mu_1 \dots \mu_{n-1}} = \sqrt{|g|} V^\mu \varepsilon_{\mu\mu_1 \dots \mu_{n-1}}.$$

Furthermore

$$\begin{aligned} d\omega &= \frac{1}{(n-1)!} \left(\partial_{\mu_1} \omega_{\mu_2 \dots \mu_n}\right) dx^{\mu_1} \wedge dx^{\mu_2} \wedge \dots \wedge dx^{\mu_n} \\ &= \frac{1}{(n-1)!} \partial_{\mu_1} \left(\sqrt{|g|} V^\mu \varepsilon_{\mu\mu_2 \dots \mu_n}\right) dx^{\mu_1} \wedge dx^{\mu_2} \wedge \dots \wedge dx^{\mu_n} \\ &= \frac{1}{n!} \partial_\mu \left(\sqrt{|g|} V^\mu \varepsilon_{\mu_1\mu_2 \dots \mu_n}\right) dx^{\mu_1} \wedge dx^{\mu_2} \wedge \dots \wedge dx^{\mu_n}. \end{aligned}$$

In the second line we must have $\mu = \mu_1$ due to the presence of $\varepsilon_{\mu\mu_2 \dots \mu_n}$ and $dx^{\mu_1} \wedge dx^{\mu_2} \wedge \dots \wedge dx^{\mu_n}$. We may therefore exchange the two covariant indices μ and μ_1. After swapping the two covariant indices we sum without the restriction $\mu = \mu_1$ over all pairs of indices (μ, μ_1). This overcounts each term n times, which is compensated by an additional factor $1/n$.

Furthermore we have $\partial_\mu \varepsilon_{\mu_1 \mu_2 \dots \mu_n} = 0$ and we obtain therefore

$$d\omega = \partial_\mu \left(\sqrt{|g|} V^\mu \right) dx^1 \wedge \dots \wedge dx^n.$$

For the Levi-Civita connection we have

$$\nabla_\mu V^\mu = \partial_\mu V^\mu + \Gamma^\mu{}_{\mu\nu} V^\nu = \frac{1}{\sqrt{|g|}} \partial_\mu \left(\sqrt{|g|} V^\mu \right).$$

Here we used

$$\Gamma^\mu{}_{\mu\nu} = \frac{1}{\sqrt{|g|}} \partial_\nu \sqrt{|g|}.$$

We obtain

$$d\omega = \left(\nabla_\mu V^\mu \right) \sqrt{|g|} d^n x$$

and hence the left-hand side of Stokes' theorem is equivalent to

$$\int\limits_M d\omega = \int\limits_M d^n x \sqrt{|g|} \nabla_\mu V^\mu.$$

Let us now consider the right-hand side of Stokes' theorem, which includes the integration over the boundary of M. The boundary ∂M is a $(n-1)$-dimensional hypersurface. It is convenient to use Gaussian normal coordinates $(z, y_1, \dots, y_{n-1})$, where the coordinates $(y_1, \dots, y_{n-1})$ parametrise the $(n-1)$-dimensional hypersurface ∂M and z is a coordinate parametrising the normal direction given by the normal vector n_μ. The induced metric on ∂M is given by

$$\gamma_{\alpha\beta} = \frac{\partial x^\mu}{\partial y^\alpha} \frac{\partial x^\nu}{\partial y^\beta} g_{\mu\nu}.$$

We may express the full metric g on M in terms of the Gaussian normal coordinates:

$$g = \pm dz \otimes dz + \gamma_{\alpha\beta} dy^\alpha \otimes dy^\beta,$$

i.e. there are no mixed terms $dy^\alpha \otimes dz$. In these coordinates we have

$$\sqrt{|g|} = \sqrt{|\gamma|}.$$

The volume element on the boundary is

$$\sqrt{|\gamma|} dy^1 \wedge \dots \wedge dy^{n-1}.$$

With the help of the unit normal vector n^μ we may write the volume element on the boundary ∂M in a coordinate-independent way:

$$\frac{1}{(n-1)!} \sqrt{|g|} n^{\mu_1} \varepsilon_{\mu_1 \mu_2 \dots \mu_n} dx^{\mu_2} \wedge \dots \wedge dx^{\mu_n}.$$

For the right-hand side of Stokes' theorem we obtain therefore

$$\begin{aligned}
\int\limits_{\partial M} \omega &= \int\limits_{\partial M} \frac{1}{(n-1)!} \sqrt{|g|} V^\mu \varepsilon_{\mu\mu_2\ldots\mu_n} dx^{\mu_2} \wedge \ldots dx^{\mu_n} \\
&= \int\limits_{\partial M} \frac{1}{(n-1)!} \sqrt{|g|} V^\mu n_\mu n^{\mu_1} \varepsilon_{\mu_1\mu_2\ldots\mu_n} dx^{\mu_2} \wedge \ldots dx^{\mu_n} \\
&= \int\limits_{\partial M} d^{n-1}y \sqrt{|\gamma|} V^\mu n_\mu.
\end{aligned}$$

It remains to discuss the sign of the unit normal vector n_μ. From the original formulation of Stokes' theorem it follows that the covariant unit normal vector n_μ is outward-pointing.

Please note that on a Lorentzian manifold the contravariant unit normal vector n^μ points outwards, if n^μ is time-like, but points inwards, if n^μ is space-like. On a Riemannian manifold, the contravariant unit normal vector n^μ is always outward-pointing.

5.9 The curvature tensor

Preliminary remark: Let

$$X = X^\mu e_\mu = X^\mu \frac{\partial}{\partial x^\mu}$$

be a vector field. A vector field acts on a functions as a directional derivative:

$$X(f) = X^\mu \frac{\partial}{\partial x^\mu} f.$$

Let

$$Y = Y^\nu \frac{\partial}{\partial x^\nu}$$

be a further vector field. We define the **Lie bracket** $[X,Y]$ as

$$[X,Y](f) = X(Y(f)) - Y(X(f)).$$

We have

$$\begin{aligned}
X(Y(f)) &= X^\mu \partial_\mu (Y^\nu \partial_\nu f) = X^\mu \left(\partial_\mu Y^\nu\right)(\partial_\nu f) + X^\mu Y^\nu \partial_\mu \partial_\nu f, \\
Y(X(f)) &= Y^\mu \partial_\mu (X^\nu \partial_\nu f) = Y^\mu \left(\partial_\mu X^\nu\right)(\partial_\nu f) + Y^\mu X^\nu \partial_\mu \partial_\nu f,
\end{aligned}$$

and hence

$$[X,Y](f) = \left(X^\mu \partial_\mu Y^\nu - Y^\mu \partial_\mu X^\nu\right) \partial_\nu f.$$

The Lie bracket is again a vector field. The components of this vector field are given by

$$[X,Y] = \left(X^\mu \partial_\mu Y^\nu - Y^\mu \partial_\mu X^\nu\right) e_\nu.$$

Remark: Neither XY nor YX are vector fields, since both contain second derivatives. The second derivatives cancel in the combination $[X,Y]$. Since only first derivatives remain, the combination $[X,Y]$is again a vector field.

Remark: An important special case is given by

$$\left[e_\mu, e_\nu\right] = 0.$$

(This is most easily seen by letting $e_\mu = \sum X^\sigma e_\sigma$ with $X^\sigma = 0$ for $\mu \neq \sigma$ and $X^\sigma = 1$ for $\mu = \sigma$.)

Since the connection coefficients $C^\lambda_{\mu\nu}$ do not transform as a tensor, they cannot have any intrinsic meaning as a measure of the curvature of a manifold. As intrinsic objects we have the **torsion tensor**

$$\begin{aligned} T &: \mathrm{Vect}(M) \otimes \mathrm{Vect}(M) \to \mathrm{Vect}(M) \\ & \quad T(X,Y) = \nabla_X Y - \nabla_Y X - [X,Y] \end{aligned}$$

and **Riemann's curvature tensor**

$$\begin{aligned} R &: \mathrm{Vect}(M) \otimes \mathrm{Vect}(M) \otimes \mathrm{Vect}(M) \to \mathrm{Vect}(M) \\ & \quad R(X,Y,Z) = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z. \end{aligned}$$

Obviously, R and T are anti-symmetric in X and Y:

$$\begin{aligned} T(X,Y) &= -T(Y,X) \\ R(X,Y,Z) &= -R(Y,X,Z) \end{aligned}$$

Using the coordinate representation we have

$$\begin{aligned} T(e_\mu, e_\nu) &= T^\lambda_{\mu\nu} e_\lambda, \\ R(e_\mu, e_\nu, e_\lambda) &= R^\kappa_{\lambda\mu\nu} e_\kappa. \end{aligned}$$

Remark: Note the position of the index λ!

With the help of

$$\nabla_\mu e_\nu = C^\lambda_{\mu\nu} e_\lambda, \qquad \left[e_\mu, e_\nu\right] = 0,$$

we determine $T^\lambda_{\mu\nu}$ and $R^\kappa_{\lambda\mu\nu}$:

$$\begin{aligned}
T(e_\mu, e_\nu) &= \nabla_\mu e_\nu - \nabla_\nu e_\mu - [e_\mu, e_\nu] \\
&= C^\lambda_{\mu\nu} e_\lambda - C^\lambda_{\nu\mu} e_\lambda \\
&= \left(C^\lambda_{\mu\nu} - C^\lambda_{\nu\mu}\right) e_\lambda, \\
R(e_\mu, e_\nu, e_\lambda) &= \nabla_\mu \nabla_\nu e_\lambda - \nabla_\nu \nabla_\mu e_\lambda - \nabla_{[e_\mu, e_\nu]} e_\lambda \\
&= \nabla_\mu C^\kappa_{\nu\lambda} e_\kappa - \nabla_\nu C^\kappa_{\mu\lambda} e_\kappa \\
&= \left(\nabla_\mu C^\kappa_{\nu\lambda}\right) e_\kappa + C^\kappa_{\nu\lambda} \nabla_\mu e_\kappa - \left(\nabla_\nu C^\kappa_{\mu\lambda}\right) e_\kappa - C^\kappa_{\mu\lambda} \nabla_\nu e_\kappa \\
&= \left(\partial_\mu C^\kappa_{\nu\lambda}\right) e_\kappa + C^\kappa_{\nu\lambda} C^\eta_{\mu\kappa} e_\eta - \left(\partial_\nu C^\kappa_{\mu\lambda}\right) e_\kappa - C^\kappa_{\mu\lambda} C^\eta_{\nu\kappa} e_\eta \\
&= \left(\partial_\mu C^\kappa_{\nu\lambda} - \partial_\nu C^\kappa_{\mu\lambda} + C^\eta_{\nu\lambda} C^\kappa_{\mu\eta} - C^\eta_{\mu\lambda} C^\kappa_{\nu\eta}\right) e_\kappa.
\end{aligned}$$

In summary we have

$$\begin{aligned}
T^\lambda_{\mu\nu} &= C^\lambda_{\mu\nu} - C^\lambda_{\nu\mu}, \\
R^\kappa_{\lambda\mu\nu} &= \partial_\mu C^\kappa_{\nu\lambda} - \partial_\nu C^\kappa_{\mu\lambda} + C^\eta_{\nu\lambda} C^\kappa_{\mu\eta} - C^\eta_{\mu\lambda} C^\kappa_{\nu\eta}.
\end{aligned}$$

Let us now specialise to the Levi-Civita connection. In this case the torsion tensor vanishes and the connection coefficients $C^\kappa_{\mu\nu}$ equal the Christoffel symbols $\Gamma^\kappa_{\mu\nu}$:

$$C^\kappa_{\mu\nu} = \Gamma^\kappa_{\mu\nu} = \frac{1}{2} g^{\kappa\lambda} \left(\partial_\mu g_{\nu\lambda} + \partial_\nu g_{\mu\lambda} - \partial_\lambda g_{\mu\nu}\right).$$

In this case we may express Riemann's curvature tensor through the Christoffel symbols:

$$R^\kappa_{\lambda\mu\nu} = \partial_\mu \Gamma^\kappa_{\nu\lambda} - \partial_\nu \Gamma^\kappa_{\mu\lambda} + \Gamma^\eta_{\nu\lambda} \Gamma^\kappa_{\mu\eta} - \Gamma^\eta_{\mu\lambda} \Gamma^\kappa_{\nu\eta}.$$

Remark: For $R_{\kappa\lambda\mu\nu} = g_{\kappa\rho} R^\rho_{\lambda\mu\nu}$ we find

$$R_{\kappa\lambda\mu\nu} = \frac{1}{2}\left(\frac{\partial^2 g_{\kappa\nu}}{\partial x^\lambda \partial x^\mu} - \frac{\partial^2 g_{\lambda\nu}}{\partial x^\kappa \partial x^\mu} + \frac{\partial^2 g_{\lambda\mu}}{\partial x^\kappa \partial x^\nu} - \frac{\partial^2 g_{\kappa\mu}}{\partial x^\lambda \partial x^\nu}\right) + g_{\xi\eta}\left(\Gamma^\xi_{\kappa\nu} \Gamma^\eta_{\lambda\mu} - \Gamma^\xi_{\kappa\mu} \Gamma^\eta_{\lambda\nu}\right).$$

The tensor $R_{\kappa\lambda\mu\nu}$ has the following symmetries:

$$\begin{aligned}
R_{\kappa\lambda\mu\nu} &= -R_{\kappa\lambda\nu\mu}, \\
R_{\kappa\lambda\mu\nu} &= -R_{\lambda\kappa\mu\nu}, \\
R_{\kappa\lambda\mu\nu} &= R_{\mu\nu\kappa\lambda}.
\end{aligned}$$

The **Ricci tensor** is defined as the following contraction of the curvature tensor:

$$Ric_{\mu\nu} = R^\lambda_{\mu\lambda\nu}$$

The Ricci tensor is symmetric:

$$Ric_{\mu\nu} = Ric_{\nu\mu}.$$

The **scalar curvature** is defined by

$$R = g^{\mu\nu} Ric_{\mu\nu}$$

As **Einstein tensor** we denote the following combination:

$$G_{\mu\nu} = Ric_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R$$

Bianchi identities:

$$\begin{aligned} R_{\kappa\lambda\mu\nu} + R_{\kappa\mu\nu\lambda} + R_{\kappa\nu\lambda\mu} &= 0, \\ \nabla_\rho R_{\kappa\lambda\mu\nu} + \nabla_\kappa R_{\lambda\rho\mu\nu} + \nabla_\lambda R_{\rho\kappa\mu\nu} &= 0. \end{aligned}$$

Proof of the first Bianchi identity: Let us first note two equivalent formulations of Bianchi's first identity:

$$\begin{aligned} R^\kappa_{\ \lambda\mu\nu} + R^\kappa_{\ \mu\nu\lambda} + R^\kappa_{\ \nu\lambda\mu} &= 0, \\ R(X,Y,Z) + R(Y,Z,X) + R(Z,X,Y) &= 0. \end{aligned}$$

In order to prove the first Bianchi identity we start from the vanishing of the torsion tensor:

$$T(X,Y) = \nabla_X Y - \nabla_Y X - [X,Y] = 0.$$

Taking the covariant derivative, we obtain

$$\begin{aligned} \nabla_Z (\nabla_X Y - \nabla_Y X - [X,Y]) &= 0, \\ \nabla_Z \nabla_X Y - \nabla_Z \nabla_Y X - \nabla_Z [X,Y] &= 0. \end{aligned}$$

We focus on the term $\nabla_Z [X,Y]$ and use again the condition that the torsion tensor vanishes:

$$\nabla_Z [X,Y] - \nabla_{[X,Y]} Z - [Z,[X,Y]] = 0.$$

Thus we obtain

$$\nabla_Z \nabla_X Y - \nabla_Z \nabla_Y X - \nabla_{[X,Y]} Z - [Z,[X,Y]] = 0.$$

If we now sum over the three cyclic permutations of (X,Y,Z) and by using the Jacobi identity

$$[Z,[X,Y]] + [X,[Y,Z]] + [Y,[Z,X]] = 0.$$

we obtain

$$\begin{aligned} & \nabla_Z \nabla_X Y - \nabla_Z \nabla_Y X - \nabla_{[X,Y]} Z \\ & + \nabla_X \nabla_Y Z - \nabla_X \nabla_Z Y - \nabla_{[Y,Z]} X \\ & + \nabla_Y \nabla_Z X - \nabla_Y \nabla_X Z - \nabla_{[Z,X]} Y = 0, \end{aligned}$$

or written in a slightly different way

$$R(X,Y,Z)+R(Y,Z,X)+R(Z,X,Y) \;=\; 0.$$

Proof of the second Bianchi identity: Equivalent formulations of Bianchi's second identity are

$$\begin{aligned}
\nabla_\rho R_{\mu\nu\kappa\lambda}+\nabla_\kappa R_{\mu\nu\lambda\rho}+\nabla_\lambda R_{\mu\nu\rho\kappa} &= 0,\\
\nabla_\rho R^{\mu}_{\ \nu\kappa\lambda}+\nabla_\kappa R^{\mu}_{\ \nu\lambda\rho}+\nabla_\lambda R^{\mu}_{\ \nu\rho\kappa} &= 0,\\
(\nabla_X R)(Y,Z,V)+(\nabla_Y R)(Z,X,V)+(\nabla_Z R)(X,Y,V) &= 0.
\end{aligned}$$

In the second line please note that the metric is covariantly constant for the Levi-Civita connection ($\nabla_\rho g^{\kappa\mu}=0$). Hence we may exchange the covariant derivative with the raising of indices. In order to prove Bianchi's second identity we introduce the following notation: Let S be the operation, which sums over the three cyclic permutations of (X,Y,Z). With this notation we have to show

$$S\,(\nabla_Z R)\,(X,Y,V) \;=\; 0.$$

We start again with the vanishing of the torsion tensor $T(X,Y)=0$ and obtain

$$R\,(T(X,Y),Z,V) \;=\; R\,(\nabla_X Y,Z,V)-R\,(\nabla_Y X,Z,V)-R\,([X,Y]\,,Z,V)=0.$$

Summation over the three cyclic permutation of (X,Y,Z) gives

$$S\,(R\,(\nabla_Z X,Y,V)-R\,(\nabla_Z Y,X,V)-R\,([X,Y]\,,Z,V)) \;=\; 0,$$

and since Riemann's curvature tensor is anti-symmetric in the first two arguments:

$$S\,(R\,(\nabla_Z X,Y,V)+R\,(X,\nabla_Z Y,V)-R\,([X,Y]\,,Z,V)) \;=\; 0,$$

We now consider

$$\nabla_Z\,(R(X,Y,V)) \;=\; (\nabla_Z R)\,(X,Y,V)+R\,(\nabla_Z X,Y,V)+R\,(X,\nabla_Z Y,V)+R\,(X,Y,\nabla_Z V)\,.$$

Using the above relation we obtain after symmetrisation

$$S\,(\nabla_Z\,(R(X,Y,V))-(\nabla_Z R)\,(X,Y,V)-R\,(X,Y,\nabla_Z V)-R\,([X,Y]\,,Z,V)) \;=\; 0.$$

We are going to prove

$$S\,[\nabla_Z\,(R(X,Y,V))-R\,(X,Y,\nabla_Z V)-R\,([X,Y]\,,Z,V)] \;=\; 0,$$

this will then imply Bianchi's second identity

$$S\,(\nabla_Z R)\,(X,Y,V) \;=\; 0.$$

We have

$$
\begin{aligned}
&\nabla_Z(R(X,Y,V)) - R(X,Y,\nabla_Z V) - R([X,Y],Z,V) = \\
&= \left(\nabla_Z\nabla_X\nabla_Y - \nabla_Z\nabla_Y\nabla_X - \nabla_Z\nabla_{[X,Y]}\right)V \\
&\quad - \left(\nabla_X\nabla_Y\nabla_Z - \nabla_Y\nabla_X\nabla_Z - \nabla_{[X,Y]}\nabla_Z\right)V \\
&\quad - \left(\nabla_{[X,Y]}\nabla_Z - \nabla_Z\nabla_{[X,Y]} - \nabla_{[[X,Y],Z]}\right)V \\
&= [\nabla_Z,[\nabla_X,\nabla_Y]]V + \nabla_{[[X,Y],Z]}V.
\end{aligned}
$$

If we now sum over the three cyclic permutations of (X,Y,Z) we have due to the Jacobi identity

$$S\left([\nabla_Z,[\nabla_X,\nabla_Y]]V + \nabla_{[[X,Y],Z]}V\right) = 0.$$

This completes the proof of Bianchi's second identity.

An important corollary of Bianchi's second identity is obtained through the following steps: Contracting the indices κ and μ in Bianchi's second identity we obtain:

$$g^{\mu\kappa}\nabla_\rho R_{\kappa\lambda\mu\nu} + \nabla^\mu R_{\lambda\rho\mu\nu} + g^{\mu\kappa}\nabla_\lambda R_{\rho\kappa\mu\nu} = 0.$$

For the Levi-Civita connection the metric is covariantly constant $\nabla_\rho g^{\kappa\mu} = 0$ and we may exchange contraction and covariant derivative:

$$\nabla_\rho Ric_{\lambda\nu} + \nabla^\mu R_{\lambda\rho\mu\nu} - \nabla_\lambda Ric_{\rho\nu} = 0.$$

If we further contract λ and ν, we obtain

$$
\begin{aligned}
\nabla_\rho R - \nabla^\mu Ric_{\rho\mu} - \nabla^\nu Ric_{\rho\nu} &= 0. \\
\nabla_\rho R - 2\nabla^\mu Ric_{\rho\mu} &= 0, \\
-2\nabla^\mu\left(Ric_{\mu\rho} - \frac{1}{2}g_{\mu\rho}R\right) &= 0.
\end{aligned}
$$

Expressed differently, we obtain

$$\nabla^\mu G_{\mu\nu} = 0.$$

5.10 Symmetries and Killing vectors

Symmetries play an important role in physics. We will now discuss the concept of symmetries in the context of semi-Riemannian manifolds. For example, the Poincaré group, consisting of Lorentz transformations and translations, is the symmetry group of flat Minkowski space. Under a Poincaré transformation the coordinates transform as

$$x'^\mu = \Lambda^\mu{}_\nu x^\nu + b^\mu.$$

The metric

$$g_{\mu\nu} dx^\mu dx^\nu$$

is invariant under these transformations. Symmetries which leave the metric invariant are called **isometries**.

Let us now define isometries (i.e. symmetries, which leave the metric invariant) for an arbitrary semi-Riemannian manifold M: Let

$$f : M \to M$$

be a diffeomorphism. We call f an isometry, if

$$f^* g = g.$$

This means that for $X, Y \in T_p M$ we have

$$g_{f(p)}(f_* X, f_* Y) = g_p(X, Y).$$

The identity map, the composition of isometries and the inverse of an isometry are again isometries. The isometries form a group. Isometries conserve the length of a vector.

Example: For Minkowski space the group of isometries is given by the Poincaré group.

Killing vector fields: Let (M, g) be a semi-Riemannian manifold and $X \in \text{Vect}(M)$ a vector field on M. The vector field X is called a Killing vector field if the transformation

$$x'^\mu = x^\mu + \varepsilon X^\mu,$$

where ε is an infinitesimal quantity, is an isometry. In this case we have

$$\frac{\partial(x^\kappa + \varepsilon X^\kappa)}{\partial x^\mu} \frac{\partial(x^\lambda + \varepsilon X^\lambda)}{\partial x^\nu} g_{\kappa\lambda}(x + \varepsilon X) = g_{\mu\nu}(x).$$

With

$$g_{\kappa\lambda}(x + \varepsilon X) = g_{\kappa\lambda}(x) + \varepsilon X^\sigma \partial_\sigma g_{\kappa\lambda}(x) + \mathcal{O}(\varepsilon^2)$$

we obtain

$$X^\sigma \partial_\sigma g_{\mu\nu} + g_{\kappa\nu} \partial_\mu X^\kappa + g_{\mu\lambda} \partial_\nu X^\lambda = 0.$$

This is Killing's equation. For the Levi-Civita connection we may re-write this equation as follows:

$$\nabla_\mu X_\nu + \nabla_\nu X_\mu = 0.$$

A set of Killing vector fields is called linearly dependent, if a vector field from this set can be written as a linear combination of the other vector fields with constant coefficients.

Remark: The number linearly independent Killing vector fields can be larger than the dimension of the manifold.

Example: We consider Minkowski space. The connection coefficients of the Levi-Civita connection vanish and Killing's equation reduces to

$$\partial_\mu X_\nu + \partial_\nu X_\mu = 0.$$

Obviously, the four constant vector fields

$$X^\mu_{(i)} = \delta^\mu_i, \quad 0 \le i \le 3,$$

satisfy this equation. But so do in addition the vector fields

$$X^\mu = a^{\mu\nu} x_\nu,$$

where $a^{\mu\nu}$ is anti-symmetric and constant. We therefore have $4+6=10$ linearly independent Killing vector fields, which of course correspond to the translations and the Lorentz transformations.

In an D-dimensional Euclidean space (or in an D-dimensional Minkowski space) we have

$$\frac{D(D+1)}{2}$$

linearly independent Killing vector fields, which correspond to D translations and $D(D-1)/2$ rotations (or Lorentz transformations).

In general we call a semi-Riemannian manifold (M,g) of dimension D a **maximally symmetric space**, if the number of linearly independent Killing vector fields is

$$\frac{D(D+1)}{2}.$$

In maximally symmetric space the curvature is the same at every point and in every direction, since the Killing vector fields provide D symmetries with respect to translations and $D(D-1)/2$ symmetries with respect to rotations. We may therefore try to construct the curvature tensor from tensors, which are invariant under these transformations. We have the metric and the total anti-symmetric tensor at our disposal. If we take into account the symmetry properties of Riemann's curvature tensor, we are left with a single possibility for the tensor structure:

$$R_{\kappa\lambda\mu\nu} = c\left(g_{\kappa\mu}g_{\lambda\nu} - g_{\kappa\nu}g_{\lambda\mu}\right).$$

The constant of proportionality is determined by contracting with $g^{\kappa\mu}$ and $g^{\lambda\nu}$:

$$R = c\left(D^2 - D\right),$$

and hence

$$R_{\kappa\lambda\mu\nu} = \frac{R}{D(D-1)}\left(g_{\kappa\mu}g_{\lambda\nu} - g_{\kappa\nu}g_{\lambda\mu}\right).$$

The curvature of a maximally symmetric space is fully specified by the scalar curvature R. As the curvature is the same at any point in a maximally symmetric space, the scalar curvature R is a constant in a maximally symmetric space. We distinguish the cases $R = 0$, $R > 0$ and $R < 0$.

The maximally symmetric spaces with a metric with Euclidean signature are:

$$\begin{array}{ll} R > 0 & \text{sphere } S^n, \\ R = 0 & \text{Euclidean space } \mathbb{R}^n, \\ R < 0 & \text{hyperbolic space } H^n. \end{array}$$

The maximally symmetric spaces with a metric with Lorentzian signature are:

$$\begin{array}{ll} R > 0 & \text{anti-de Sitter space } AdS^n, \\ R = 0 & \text{Minkowski space } M^n, \\ R < 0 & \text{de Sitter space } dS^n. \end{array}$$

We recall that we use the convention that a Lorentzian metric has one positive and $(n-1)$ negative eigenvalues. One finds in the literature also the opposite convention, where a Lorentzian metric has one negative and $(n-1)$ positive eigenvalues. We may obtain one case from the other case through the substitution

$$g_{\mu\nu} \rightarrow -g_{\mu\nu}.$$

Under this transformation we have

$$\begin{aligned} \Gamma^{\kappa}_{\mu\nu} &\rightarrow \Gamma^{\kappa}_{\mu\nu}, \\ R^{\kappa}_{\lambda\mu\nu} &\rightarrow R^{\kappa}_{\lambda\mu\nu}, \\ Ric_{\mu\nu} &\rightarrow Ric_{\mu\nu}, \\ R &\rightarrow -R. \end{aligned}$$

5.11 The Weyl tensor

The Ricci tensor and the scalar curvature project out the information related to traces of the Riemann curvature tensor. The trace-free part is lost. The trace-free part is captured by the Weyl tensor. The **Weyl tensor** is defined in D dimensions by

$$\begin{aligned} C_{\kappa\lambda\mu\nu} = {} & R_{\kappa\lambda\mu\nu} - \frac{2}{D-2}\left(g_{\kappa\mu}Ric_{\nu\lambda} - g_{\kappa\nu}Ric_{\mu\lambda} - g_{\lambda\mu}Ric_{\nu\kappa} + g_{\lambda\nu}Ric_{\mu\kappa}\right) \\ & + \frac{2}{(D-1)(D-2)}\left(g_{\kappa\mu}g_{\nu\lambda} - g_{\kappa\nu}g_{\mu\lambda}\right)R. \end{aligned}$$

The Weyl tensor is only defined for manifolds of dimension $D \geq 3$. For $D = 3$ the Weyl tensor vanishes identically. The Weyl tensor has the same symmetries as the Riemann curvature tensor:

$$\begin{aligned}
C_{\kappa\lambda\mu\nu} &= -C_{\kappa\lambda\nu\mu}, \\
C_{\kappa\lambda\mu\nu} &= -C_{\lambda\kappa\mu\nu}, \\
C_{\kappa\lambda\mu\nu} &= C_{\mu\nu\kappa\lambda}, \\
C_{\kappa\lambda\mu\nu} + C_{\kappa\mu\nu\lambda} + C_{\kappa\nu\lambda\mu} &= 0,
\end{aligned}$$

The Weyl tensor is also known as **conformal tensor**. The reason is as follows: Consider two metrics $g_{\mu\nu}$ and

$$g'_{\mu\nu} = \omega^2(x) g_{\mu\nu},$$

where $\omega(x)$ is an arbitrary non-vanishing function on the manifold. One finds

$$C^{\kappa}_{\lambda\mu\nu} = C'^{\kappa}_{\lambda\mu\nu}.$$

6 Einstein's equations

6.1 Relevant scales

Let us first look at the order of magnitude of the gravitational force in comparison to the electromagnetic force: The gravitational force between a proton and an anti-proton is given by

$$F_G = -\frac{Gm_p^2}{r^2}\hat{r},$$

where G denotes Newton's constant. The numerical value is

$$G = (6.67259 \pm 0.00085) \cdot 10^{-11}\ \mathrm{m}^3\mathrm{kg}^{-1}\mathrm{s}^{-2}.$$

Let's compare this to the electric force. The Coulomb force is given by

$$F_C = -\frac{1}{4\pi\varepsilon_0}\frac{e^2}{r^2}\hat{r}.$$

For the ratio of the two forces we have

$$\left|\frac{F_G}{F_C}\right| = \frac{4\pi\varepsilon_0 G m_p^2}{e^2} = 0.81 \cdot 10^{-36}.$$

The gravitational force is the weakest among the known fundamental forces (gravitational force, electromagnetic force, weak force, strong force).

Remark: The gravitational force is always attractive, contrary to the electric force, which can be attractive or repulsive.

Dimensionless quantities:

$$\alpha = \frac{1}{4\pi\varepsilon_0}\frac{e^2}{\hbar c} = 0.0072973 = \frac{1}{137.036},$$
$$\alpha_G = \frac{Gm_p^2}{\hbar c} = 5.9 \cdot 10^{-39}.$$

Planck mass:

$$M_{Pl} = \sqrt{\frac{\hbar c}{G}} = 1.221 \cdot 10^{19}\ \mathrm{GeV} = 2.177 \cdot 10^{-8}\ \mathrm{kg}.$$

The Planck mass is significantly larger than the masses of the elementary particles known today.

Planck length:

$$\lambda_{Pl} = \frac{2\pi\hbar c}{M_{Pl}c^2} = (2\pi)1.62 \cdot 10^{-35}\ \mathrm{m}.$$

The Planck length is significantly smaller than the typical range of sub-atomic forces ($\approx 10^{-18}$ m).

6.2 The equivalence principle

The equivalence principle: Let us first consider a particle in a gravitational field within non-relativistic mechanics. The Lagrange function is given by

$$L = \frac{1}{2} m_T v^2 - m_S \phi,$$

where m_T denotes the inertial mass of the particle and m_S denotes the gravitational mass of the particle. The equation of motion reads:

$$m_T \frac{d}{dt} \vec{v} = -m_S \vec{\nabla} \phi.$$

All experimental data is compatible with $m_T = m_S$. This is the weak formulation of the equivalence principle: **The gravitational mass equals the inertial mass.** Therefore:

$$\frac{d}{dt} \vec{v} = -\vec{\nabla} \phi.$$

Let us now consider a number of test particles in a homogeneous and time-independent gravitational field. In an inertial system K the equations of motion read

$$m_i \frac{d^2}{dt^2} \vec{x}^{(i)} = m_i \vec{g} + \sum_{j \neq i} \vec{F}_{ij}.$$

Let us now change from the inertial system K to a non-inertial system K', which is obtained from K by a constant acceleration $\vec{g}$, i.e.

$$\vec{y} = \vec{x} - \frac{1}{2} \vec{g} t^2.$$

In the system K' the equations of motion read

$$m_i \frac{d^2}{dt^2} \vec{y}^{(i)} = \sum_{j \neq i} \vec{F}_{ij}.$$

Strong version of the equivalence principle: For each point x of the space-time M there exists a local inertial system such that in a sufficiently small neighbourhood $U \subset M$ of x the equations of motion take the form as in special relativity. This implies that the existence of a gravitational field cannot be detected by local experiments alone.

Remark: The weak version of the equivalence principle refers only to the equation of motions for freely falling bodies, the strong version refers to all physical phenomena.

In the following we will denote by

$$\eta_{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

the known metric of flat Minkowski space-time. Within general relativity the metric will be promoted to a coordinate-dependent object. A mathematical precise formulation of the strong equivalence principle reads: For each point x_0 of space-time there exists a coordinate system such that

$$\begin{aligned} g_{\mu\nu}(x_0) &= \eta_{\mu\nu}, \\ \left. \frac{\partial g_{\mu\nu}(x)}{\partial x^\alpha} \right|_{x_0} &= 0. \end{aligned}$$

Such coordinates are called **Gauß coordinates** or **normal coordinates**.

6.3 Motion of particles in a gravitational field

Let us first consider the motion of a free particle (i.e. no forces are exerted on the particle) on a given manifold.

We recall that within Newtonian mechanics a free particle moves with constant velocity along straight lines.

Within special relativity we have the law that the motion of a free particle is a motion with constant four-velocity:

$$\frac{d}{ds} u^\mu = 0.$$

This equation of motion can be deduced with the help of the principle of least action from the action of a free particle

$$S = -mc \int\limits_a^b ds.$$

The action is proportional to the length of the path between the space-time points a and b. A minimum is obtained for the shortest path between a and b. Paths, which give the shortest path between two points are called **geodesics**.

This gives us the proper generalisation to curved manifolds: The motion of a free particle on an arbitrary semi-Riemannian manifold is given by a geodesic. For semi-Riemannian manifolds with the Levi-Civita connection there is an alternative definition for a geodesic: A geodesic is a curve along which the tangent vector is parallel transported.

Let $x^\mu(\lambda)$ be a curve and let $T^{\mu_1 \dots \mu_k}{}_{\nu_1 \dots \nu_l}$ be a tensor. The tangent vector of the curve at the point $x^\mu(0)$ is given by

$$V = \frac{dx^\mu}{d\lambda} e_\mu.$$

By definition, the tensor is parallel transported along the curve if

$$\nabla_V T^{\mu_1\dots\mu_k}{}_{\nu_1\dots\nu_l} = \frac{dx^\tau}{d\lambda}\nabla_\tau T^{\mu_1\dots\mu_k}{}_{\nu_1\dots\nu_l} = 0.$$

For a vector field (i.e. a $(1,0)$-tensor field) this equation simplifies to

$$\frac{dx^\tau}{d\lambda}\nabla_\tau V^\mu = \frac{dx^\tau}{d\lambda}\left(\partial_\tau V^\mu + \Gamma^\mu_{\tau\sigma}V^\sigma\right) = 0.$$

If we plug in for V^μ the expression for the tangent vector $V^\mu = dx^\mu/d\lambda$, we find

$$\frac{dx^\tau}{d\lambda}\left(\partial_\tau \frac{dx^\mu}{d\lambda} + \Gamma^\mu_{\tau\sigma}\frac{dx^\sigma}{d\lambda}\right) = \frac{d^2x^\mu}{d\lambda^2} + \Gamma^\mu_{\tau\sigma}\frac{dx^\tau}{d\lambda}\frac{dx^\sigma}{d\lambda} = 0.$$

The equation

$$\frac{d^2x^\mu}{d\lambda^2} + \Gamma^\mu_{\tau\sigma}\frac{dx^\tau}{d\lambda}\frac{dx^\sigma}{d\lambda} = 0$$

is called the **geodesic equation**. If all connection coefficients vanish (as for example in the case of an Euclidean space or Minkowski space), the geodesic equation reduces to

$$\frac{d^2x^\mu}{d\lambda^2} = 0,$$

which corresponds to the motion of a particle with constant velocity along straight lines.

In order to derive the geodesic equation we started from the definition of a geodesic which refers to the parallel transport of the tangent vector along the geodesic curve. Let us return to the first definition, which defines geodesics as paths of shortest length between two points. We consider the functional

$$s = \int \sqrt{g_{\mu\nu}\frac{dx^\mu}{d\lambda}\frac{dx^\nu}{d\lambda}}\, d\lambda.$$

We set

$$f = g_{\mu\nu}\frac{dx^\mu}{d\lambda}\frac{dx^\nu}{d\lambda}.$$

For the variation of the functional one obtains

$$\delta s = \int \delta\sqrt{f}\, d\lambda = \frac{1}{2}\int \frac{1}{\sqrt{f}}\delta f\, d\lambda.$$

Let us choose for the curve parameter λ the proper time (more precisely $s = c\tau$). We then find

$$f = g_{\mu\nu}\frac{dx^\mu}{ds}\frac{dx^\nu}{ds} = g_{\mu\nu}u^\mu u^\nu = 1.$$

It is therefore sufficient to consider the extrema of the simpler functional

$$I = \frac{1}{2}\int f ds = \frac{1}{2}\int g_{\mu\nu}\frac{dx^\mu}{ds}\frac{dx^\nu}{ds}ds.$$

Let us now consider

$$\begin{aligned} x^\mu &\to x^\mu + \delta x^\mu, \\ g_{\mu\nu} &\to g_{\mu\nu} + \left(\partial_\sigma g_{\mu\nu}\right)\delta x^\sigma. \end{aligned}$$

Plugging this in, we obtain

$$\delta I = \frac{1}{2}\int \left[\partial_\sigma g_{\mu\nu}\frac{dx^\mu}{ds}\frac{dx^\nu}{ds}\delta x^\sigma + g_{\mu\nu}\frac{d\left(\delta x^\mu\right)}{ds}\frac{dx^\nu}{ds} + g_{\mu\nu}\frac{dx^\mu}{ds}\frac{d\left(\delta x^\nu\right)}{ds}\right] ds.$$

For the last two terms we use partial integration, as for example

$$\begin{aligned} \frac{1}{2}\int g_{\mu\nu}\frac{dx^\mu}{ds}\frac{d\left(\delta x^\nu\right)}{ds}ds &= -\frac{1}{2}\int \left[g_{\mu\nu}\frac{d^2x^\mu}{ds^2} + \frac{dg_{\mu\nu}}{ds}\frac{dx^\mu}{ds}\right]\delta x^\nu ds \\ &= -\frac{1}{2}\int \left[g_{\mu\nu}\frac{d^2x^\mu}{ds^2} + \partial_\sigma g_{\mu\nu}\frac{dx^\sigma}{ds}\frac{dx^\mu}{ds}\right]\delta x^\nu ds. \end{aligned}$$

After partial integration we obtain

$$\delta I = -\int \left[g_{\mu\sigma}\frac{d^2x^\mu}{ds^2} + \frac{1}{2}\left(\partial_\mu g_{\nu\sigma} + \partial_\nu g_{\sigma\mu} - \partial_\sigma g_{\mu\nu}\right)\frac{dx^\mu}{ds}\frac{dx^\nu}{ds}\right]\delta x^\sigma ds.$$

The vanishing of the variation implies

$$g_{\mu\sigma}\frac{d^2x^\mu}{ds^2} + \frac{1}{2}\left(\partial_\mu g_{\nu\sigma} + \partial_\nu g_{\sigma\mu} - \partial_\sigma g_{\mu\nu}\right)\frac{dx^\mu}{ds}\frac{dx^\nu}{ds} = 0.$$

Finally, we multiply by the inverse metric. We obtain

$$\frac{d^2x^\rho}{ds^2} + \frac{1}{2}g^{\rho\sigma}\left(\partial_\mu g_{\nu\sigma} + \partial_\nu g_{\sigma\mu} - \partial_\sigma g_{\mu\nu}\right)\frac{dx^\mu}{ds}\frac{dx^\nu}{ds} = 0.$$

This is exactly the geodesic equation with the Christoffel symbols as connection coefficients. This shows that the two definitions of a geodesic are equivalent for the Levi-Civita connection.

Finally, let us give a third derivation of the geodesic equation. We generalise the known relation in flat Minkowski space in a covariant way. We start from the equation of motion for a free particle in Minkowski space:

$$\frac{d}{ds}u^\mu = 0,$$

We may re-write this as

$$du^\mu = 0.$$

The generalisation to curved space reads

$$\nabla u^\mu = 0.$$

With the definition of the covariant derivative one obtains

$$du^\mu + \Gamma^\mu_{\nu\rho} u^\nu dx^\rho = 0.$$

If we now divide again by ds, we obtain

$$\frac{d^2 x^\mu}{ds^2} + \Gamma^\mu_{\nu\rho} \frac{dx^\nu}{ds} \frac{dx^\rho}{ds} = 0.$$

This is the sought-after equation of motion. The motion of the particle is determined by the quantities $\Gamma^\mu_{\nu\rho}$. Since $\frac{d^2 x^\mu}{ds^2}$ gives the four-acceleration of the particle, we may interpret the quantity

$$-m\Gamma^\mu_{\nu\rho} u^\nu u^\rho$$

as the four-force acting on particles due to the gravitational field.

6.4 Einstein's equations

In this section we will heuristically motivate Einstein's equations. In the last section we saw that the geodesic equation can be obtained from the equation of motion in flat space $du^\mu/ds = 0$ by replacing partial derivatives with covariant derivatives. In this section we will use these "rules" to obtain the field equations for gravitation. In a subsequent section we will adopt a stricter approach and derive the field equations from an action. The rules for "minimal substitution" are:

- Replace partial derivatives by covariant derivatives.
- Replace the flat metric $\eta_{\mu\nu}$ by $g_{\mu\nu}$.

Let us consider an example. In flat Minkowski space we have

$$\partial_\mu T^{\mu\nu} = 0.$$

The generalisation to curved manifolds reads

$$\nabla_\mu T^{\mu\nu} = 0.$$

Once we obtained Einstein's equations, we would also show that in the Newtonian limit they reduce to the well-known equations of classical mechanics:

$$\frac{d^2 \vec{x}}{dt^2} = -\vec{\nabla}\Phi, \qquad \Delta\Phi = \frac{4\pi G\rho}{c^2}.$$

The Newtonian limit is defined by

- All particle velocities are small compared to the speed of light.
- The gravitational field is weak, such that it can be treated as a perturbation of flat space-time.
- The gravitational field is static (i.e. time-independent).

Let us now consider a weak static gravitational field. In general, the equation of motion for a free particle reads

$$\frac{d^2x^\mu}{ds^2} + \Gamma^\mu_{\nu\rho}\frac{dx^\nu}{ds}\frac{dx^\rho}{ds} = 0.$$

The four-velocity is given by

$$u^\mu = \frac{dx^\mu}{ds} = \left(\frac{1}{\sqrt{1-\frac{v^2}{c^2}}}, \frac{\frac{\vec{v}}{c}}{\sqrt{1-\frac{v^2}{c^2}}}\right).$$

For a slow motion (i.e. $|\vec{v}| \ll c$) we have

$$\left|\frac{d\vec{x}}{ds}\right| \ll \left|\frac{dx^0}{ds}\right|.$$

In this limit the equation of motion simplifies to

$$\frac{d^2x^\mu}{ds^2} + \Gamma^\mu{}_{00}\frac{dx^0}{ds}\frac{dx^0}{ds} = 0.$$

For a static gravitational field the Christoffel symbols reduce to

$$\Gamma^\mu{}_{00} = \frac{1}{2}g^{\mu\lambda}\left(\partial_0 g_{0\lambda} + \partial_0 g_{0\lambda} - \partial_\lambda g_{00}\right) = -\frac{1}{2}g^{\mu\lambda}\partial_\lambda g_{00}.$$

Let us now set

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu},$$

with $|h_{\mu\nu}| \ll 1$. We obtain for the inverse metric $g^{\mu\nu}$ to first order

$$g^{\mu\nu} = \eta^{\mu\nu} - h^{\mu\nu},$$

with

$$h^{\mu\nu} = \eta^{\mu\sigma}\eta^{\nu\tau}h_{\sigma\tau}.$$

For $\Gamma^\mu{}_{00}$ one finds

$$\Gamma^\mu{}_{00} = -\frac{1}{2}\eta^{\mu\lambda}\partial_\lambda h_{00},$$

We substitute this result into the equation of motion

$$\frac{d^2x^\mu}{ds^2} = \frac{1}{2}\eta^{\mu\lambda}\left(\partial_\lambda h_{00}\right)\left(\frac{dx^0}{ds}\right)^2.$$

With $ds = cd\tau$ we obtain for the spatial components of the equation of motion

$$\frac{d^2x^i}{d\tau^2} = \frac{1}{2}c^2\left(\partial^i h_{00}\right)\left(\frac{dt}{d\tau}\right)^2.$$

We divide both sides by $(dt/d\tau)^2$ and obtain

$$\frac{d^2x^i}{dt^2} = \frac{1}{2}c^2\partial^i h_{00},$$

With $\vec{\nabla} = (\partial_1, \partial_2, \partial_3) = -(\partial^1, \partial^2, \partial^3)$ we have

$$\frac{d^2\vec{x}}{dt^2} = -\frac{1}{2}c^2\vec{\nabla}h_{00}.$$

Let us compare this equation with

$$\frac{d^2\vec{x}}{dt^2} = -\vec{\nabla}\Phi.$$

We deduce that the gravitational potential is given by

$$\Phi = \frac{1}{2}c^2 h_{00}.$$

Thus

$$g_{00} = \eta_{00} + h_{00} = 1 + \frac{2}{c^2}\Phi.$$

We see that a metric of the form $g_{00} = 1 + \frac{2}{c^2}\Phi$ corresponds in the Newtonian limit to Newton's law $d^2\vec{x}/dt^2 = -\vec{\nabla}\Phi$.

Let us now seek a generalisation of Poisson's law: $\Delta\Phi = 4\pi G\rho/c^2$. (We use the convention that the mass density ρ is given in units of energy per volume, therefore an extra factor of $1/c^2$ appears.) As starting point we will assume that the mass is the source of the gravitational field. In natural units ($c = 1$) we have

$$\begin{aligned} \text{mass} &= \text{rest energy} \\ &= \text{0-component of a four-vector.} \end{aligned}$$

$$\begin{aligned} \text{mass density} &= \text{energy density} \\ &= \text{00-component of a rank 2 four-tensor.} \end{aligned}$$

We therefore expect that the energy-momentum tensor $T^{\mu\nu}$ describes the source of the gravitational field. In Minkowski space energy-momentum conservation implies:

$$\partial_\mu T^{\mu\nu} = 0.$$

In general coordinates this equation reads

$$\nabla_\mu T^{\mu\nu} = 0.$$

We therefore seek an equation involving rank 2 tensors and containing $T^{\mu\nu}$.

We further know that Newton's gravitational potential satisfies the Poisson equation

$$\Delta\Phi = \frac{4\pi G\rho}{c^2},$$

and that the mass density ρ is the 00-component of the energy-momentum tensor:

$$\rho = T^{00}.$$

We further have

$$\begin{aligned} \Phi &\approx \frac{1}{2}c^2 h_{00}, \\ g_{00} &= 1 + h_{00}. \end{aligned}$$

Therefore we find

$$\Delta g_{00} = \frac{8\pi G}{c^4} T_{00}.$$

Thus we seek an equation of the form

$$\tilde{G}_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu},$$

where the tensor $\tilde{G}_{\mu\nu}$ contains the metric and its first and second derivatives.

Let us summarise: We look for a quantity $\tilde{G}_{\mu\nu}$ with the following properties:

1. $\tilde{G}_{\mu\nu}$ is a tensor;
2. $\tilde{G}_{\mu\nu}$ contains derivatives of the metric up to second order, second derivatives of the metric occur linearly, first derivatives of the metric are allowed to occur quadratically;
3. $\tilde{G}_{\mu\nu}$ is symmetric, since $T_{\mu\nu}$ is symmetric;
4. $\nabla^\mu \tilde{G}_{\mu\nu} = 0$, since $T_{\mu\nu}$ is conserved ($\nabla^\mu T_{\mu\nu} = 0$);

5. For weak static gravitational fields we have

$$\tilde{G}_{00} \rightarrow \Delta g_{00}.$$

The first two points imply, that $\tilde{G}_{\mu\nu}$ must be a linear combination of $Ric_{\mu\nu}$ and $g_{\mu\nu}R$, other tensors are not available. Hence

$$\tilde{G}_{\mu\nu} = c_1 Ric_{\mu\nu} + c_2 g_{\mu\nu} R.$$

This ansatz also satisfies condition 3. We already know that the Einstein tensor satisfies

$$\nabla_\mu G^{\mu\kappa} = 0.$$

Since $G_{\mu\nu} = Ric_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R$ we conclude

$$c_2 = -\frac{1}{2} c_1.$$

Condition 5 implies that the constant of proportionality is given by

$$c_1 = 1.$$

Hence

$$\tilde{G}_{\mu\nu} = G_{\mu\nu}$$

and Einstein's field equations read

$$\begin{aligned} G_{\mu\nu} &= \frac{8\pi G}{c^4} T_{\mu\nu}, \\ Ric_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R &= \frac{8\pi G}{c^4} T_{\mu\nu}. \end{aligned}$$

Uniqueness of Einstein's equations: Assumptions 1-4 are indispensable, but it could be possible that small deviations from Newton's law remained undetected up to today. It can be shown that Einstein's equations are unique up to an additional term

$$\Lambda g_{\mu\nu}.$$

Λ is called the **cosmological constant**. The cosmological constant was introduced by Einstein in 1917 and later discarded ("größte Eselei ..."). Today there is strong evidence that $\Lambda \neq 0$. Einstein's equations with a cosmological constant read:

$$Ric_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R - \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu},$$

Remark concerning the sign of the term $\Lambda g_{\mu\nu}$: If one uses the signature $(-,+,+,+)$ instead of the signature $(+,-,-,-)$ adopted in these lectures, the terms $Ric_{\mu\nu}$ and $g_{\mu\nu}R$ won't change

sign, however the metric $g_{\mu\nu}$ will change the sign. In order to have with both conventions the same numerical value for the cosmological constant one finds in the literature which uses the convention $(-,+,+,+)$ the expression $Ric_{\mu\nu} - 1/2 g_{\mu\nu} R + \Lambda g_{\mu\nu}$.
In the presence of a cosmological constant we obtain in the Newtonian limit

$$\Delta\Phi = \frac{4\pi G\rho}{c^2} + \frac{1}{2}c^2\Lambda.$$

We see that a non-vanishing cosmological constant Λ implies a homogeneous static energy density in the universe given by

$$\rho_{vac} = \frac{c^4}{8\pi G}\Lambda.$$

Remarks:

- Einstein's equations are non-linear differential equations. They contain second derivatives of $g_{\mu\nu}$, but also products of first derivatives and $g_{\mu\nu}$. The non-linearity implies that the superposition principle does not apply to gravity.

- We may contract Einstein's equations with $g^{\mu\nu}$. This yields

$$R - 2R - 4\Lambda = \frac{8\pi G}{c^4} T,$$

where we set $T = g^{\mu\nu} T_{\nu\mu}$. This equation can be solved for R:

$$R = -\frac{8\pi G}{c^4} T - 4\Lambda.$$

We may now substitute this expression for the scalar curvature into Einstein's field equations and obtain

$$Ric_{\mu\nu} = \frac{8\pi G}{c^4}\left(T_{\mu\nu} - \frac{1}{2} g_{\mu\nu} g^{\rho\sigma} T_{\sigma\rho}\right) - \Lambda g_{\mu\nu}.$$

- In empty space we have $T_{\mu\nu} = 0$. If in addition the cosmological constant is vanishing as well, one has $Ric_{\mu\nu} = 0$. However, this does in general not imply that $R_{\mu\nu\rho\sigma} = 0$, i.e. that the curvature tensor is vanishing. Remark: In dimensions $D = 2$ or $D = 3$ one can show that $Ric_{\mu\nu} = 0$ implies $R_{\mu\nu\rho\sigma} = 0$.

6.5 The action of general relativity

Let us first consider the gravitational field alone, i.e. without additional matter fields. The **Einstein-Hilbert action** with a cosmological constant reads:

$$S_{EH} = -\frac{c^3}{16\pi G}\int d^4x \sqrt{-g}\,(R + 2\Lambda)$$

Here we used the notation

$$g = \det g_{\mu\nu}.$$

We now derive the equations of motion through the variation of the metric. It is technically simpler to work out the variation with respect to the inverse metric $g^{\mu\nu}$ instead of the variation with respect to the metric $\delta g_{\mu\nu}$. Since

$$g^{\mu\rho} g_{\rho\nu} = \delta^{\mu}_{\nu}$$

we have

$$\begin{aligned} \left(\delta g^{\mu\rho}\right) g_{\rho\nu} + g^{\mu\rho}\left(\delta g_{\rho\nu}\right) &= 0, \\ \delta g_{\mu\nu} &= -g_{\mu\rho} g_{\nu\sigma} \delta g^{\rho\sigma}. \end{aligned}$$

With $R = g^{\mu\nu} Ric_{\mu\nu}$ we obtain three terms for the variation of the action:

$$\begin{aligned} \delta S_{EH} &= -\frac{c^3}{16\pi G} \delta \int d^4x \sqrt{-g} \left(g^{\mu\nu} Ric_{\mu\nu} + 2\Lambda\right) \\ &= -\frac{c^3}{16\pi G} \left[\underbrace{\int d^4x \sqrt{-g} g^{\mu\nu} \delta Ric_{\mu\nu}}_{(\delta S)_1} + \underbrace{\int d^4x \sqrt{-g} Ric_{\mu\nu} \delta g^{\mu\nu}}_{(\delta S)_2} + \underbrace{\int d^4x \left(g^{\mu\nu} Ric_{\mu\nu} + 2\Lambda\right) \delta\sqrt{-g}}_{(\delta S)_3} \right]. \end{aligned}$$

The second term is already in the desired form of an expression multiplied by $\delta g^{\mu\nu}$.

Let us start with the first term. We recall that the Ricci tensor is given as a contraction of Riemann's curvature tensor. The curvature tensor is expressed in turn in terms of the Christoffel symbols:

$$R^{\kappa}_{\lambda\mu\nu} = \partial_\mu \Gamma^{\kappa}_{\nu\lambda} - \partial_\nu \Gamma^{\kappa}_{\mu\lambda} + \Gamma^{\eta}_{\nu\lambda} \Gamma^{\kappa}_{\mu\eta} - \Gamma^{\eta}_{\mu\lambda} \Gamma^{\kappa}_{\nu\eta}.$$

Therefore we consider first the variation of Riemann's curvature tensor with respect to the Christoffel symbol.

$$\Gamma'^{\kappa}_{\mu\nu} = \Gamma^{\kappa}_{\mu\nu} + \delta\Gamma^{\kappa}_{\mu\nu}.$$

At this point it is important to recall that the Christoffel symbol is not a $(1,2)$-tensor! In order to find the transformation law for the Christoffel symbol under a coordinate transformation we consider

$$\nabla_{\mu'} V^{\nu'} = \frac{\partial x^{\mu}}{\partial x^{\mu'}} \frac{\partial x^{\nu'}}{\partial x^{\nu}} \nabla_\mu V^\nu.$$

The left-hand side may be expressed as

$$
\begin{aligned}
\nabla_{\mu'} V^{\nu'} &= \partial_{\mu'} V^{\nu'} + \Gamma^{\nu'}_{\mu'\lambda'} V^{\lambda'} \\
&= \frac{\partial x^\mu}{\partial x^{\mu'}} \partial_\mu \left(\frac{\partial x^{\nu'}}{\partial x^\nu} V^\nu \right) + \Gamma^{\nu'}_{\mu'\lambda'} \frac{\partial x^{\lambda'}}{\partial x^\lambda} V^\lambda \\
&= \frac{\partial x^\mu}{\partial x^{\mu'}} \frac{\partial x^{\nu'}}{\partial x^\nu} \partial_\mu V^\nu + \Gamma^{\nu'}_{\mu'\lambda'} \frac{\partial x^{\lambda'}}{\partial x^\lambda} V^\lambda + \frac{\partial x^\mu}{\partial x^{\mu'}} V^\nu \frac{\partial^2 x^{\nu'}}{\partial x^\mu \partial x^\nu}.
\end{aligned}
$$

For the right-hand side we obtain

$$
\frac{\partial x^\mu}{\partial x^{\mu'}} \frac{\partial x^{\nu'}}{\partial x^\nu} \nabla_\mu V^\nu = \frac{\partial x^\mu}{\partial x^{\mu'}} \frac{\partial x^{\nu'}}{\partial x^\nu} \partial_\mu V^\nu + \frac{\partial x^\mu}{\partial x^{\mu'}} \frac{\partial x^{\nu'}}{\partial x^\nu} \Gamma^\nu_{\mu\lambda} V^\lambda.
$$

Therefore

$$
\Gamma^{\nu'}_{\mu'\lambda'} \frac{\partial x^{\lambda'}}{\partial x^\lambda} V^\lambda + \frac{\partial x^\mu}{\partial x^{\mu'}} V^\lambda \frac{\partial^2 x^{\nu'}}{\partial x^\mu \partial x^\lambda} = \frac{\partial x^\mu}{\partial x^{\mu'}} \frac{\partial x^{\nu'}}{\partial x^\nu} \Gamma^\nu_{\mu\lambda} V^\lambda.
$$

Here we replaced in the second term on the left-hand side the summation index ν by λ. Since this has to hold for arbitrary V^λ one obtains after multiplication with $\partial x^\lambda / \partial x^{\lambda'}$

$$
\Gamma^{\nu'}_{\mu'\lambda'} = \frac{\partial x^\mu}{\partial x^{\mu'}} \frac{\partial x^{\nu'}}{\partial x^\nu} \frac{\partial x^\lambda}{\partial x^{\lambda'}} \Gamma^\nu_{\mu\lambda} - \frac{\partial x^\mu}{\partial x^{\mu'}} \frac{\partial x^\lambda}{\partial x^{\lambda'}} \frac{\partial^2 x^{\nu'}}{\partial x^\mu \partial x^\lambda}.
$$

Let $C^\nu_{\mu\lambda}$ and $\tilde{C}^\nu_{\mu\lambda}$ be two connections. The difference transforms as

$$
C^{\nu'}_{\mu'\lambda'} - \tilde{C}^{\nu'}_{\mu'\lambda'} = \frac{\partial x^\mu}{\partial x^{\mu'}} \frac{\partial x^{\nu'}}{\partial x^\nu} \frac{\partial x^\lambda}{\partial x^{\lambda'}} \left(C^\nu_{\mu\lambda} - \tilde{C}^\nu_{\mu\lambda} \right),
$$

since all terms with second derivatives cancel out. Therefore, the difference $C^\nu_{\mu\lambda} - \tilde{C}^\nu_{\mu\lambda}$ is a $(1,2)$-tensor. In particular this implies that the variation of the Christoffel symbol

$$
\delta\Gamma^\kappa_{\mu\nu} = \Gamma'^\kappa_{\mu\nu} - \Gamma^\kappa_{\mu\nu}
$$

transforms as a tensor. Hence

$$
\nabla_\mu \left(\delta\Gamma^\kappa_{\nu\lambda} \right) = \partial_\mu \left(\delta\Gamma^\kappa_{\nu\lambda} \right) + \Gamma^\kappa_{\mu\rho} \delta\Gamma^\rho_{\nu\lambda} - \Gamma^\rho_{\mu\nu} \delta\Gamma^\kappa_{\rho\lambda} - \Gamma^\rho_{\mu\lambda} \delta\Gamma^\kappa_{\nu\rho}.
$$

The variation of Riemann's curvature tensor with respect to the Christoffel symbol yields

$$
\begin{aligned}
\delta R^\kappa{}_{\lambda\mu\nu} &= \partial_\mu \delta\Gamma^\kappa_{\nu\lambda} - \partial_\nu \delta\Gamma^\kappa_{\mu\lambda} + \delta\Gamma^\eta_{\nu\lambda} \Gamma^\kappa_{\mu\eta} + \Gamma^\eta_{\nu\lambda} \delta\Gamma^\kappa_{\mu\eta} - \delta\Gamma^\eta_{\mu\lambda} \Gamma^\kappa_{\nu\eta} - \Gamma^\eta_{\mu\lambda} \delta\Gamma^\kappa_{\nu\eta} \\
&= \nabla_\mu \left(\delta\Gamma^\kappa_{\nu\lambda} \right) - \nabla_\nu \left(\delta\Gamma^\kappa_{\mu\lambda} \right).
\end{aligned}
$$

We then express the variation $\delta\Gamma^\kappa_{\mu\nu}$ in terms of the variation $\delta g^{\mu\nu}$:

$$
\delta\Gamma^\kappa_{\mu\nu} = -\frac{1}{2} \left[g_{\lambda\mu} \nabla_\nu \delta g^{\lambda\kappa} + g_{\lambda\nu} \nabla_\mu \delta g^{\lambda\kappa} - g_{\mu\alpha} g_{\nu\beta} \nabla^\kappa \delta g^{\alpha\beta} \right].
$$

Combining all ingredients, we obtain for the first term

$$(\delta S)_1 = \int d^4x \sqrt{-g} \nabla_\sigma \left[g_{\mu\nu} \nabla^\sigma (\delta g^{\mu\nu}) - \nabla_\lambda \left(\delta g^{\sigma\lambda} \right) \right].$$

This integral is a covariant divergence of a vector and can be re-written as a boundary integral at infinity. This term does not contribute to the variation.

Let us now consider $(\delta S)_3$: We have to calculate the variation of the determinant of g. In this respect, the following formula is useful: For any quadratic matrix with non-vanishing determinant we have

$$\ln(\det M) = \mathrm{Tr}\ (\ln M).$$

The logarithm of a matrix is defined by

$$\exp(\ln M) = M,$$

and the exponential function is defined by the series expansion. If $M = \mathrm{diag}(\lambda_1, ..., \lambda_n)$ is diagonal, the above formula is immediately clear:

$$\ln(\lambda_1 \cdot \lambda_2 \cdot ... \cdot \lambda_n) = \ln\lambda_1 + \ln\lambda_2 + ... + \ln\lambda_n.$$

For an arbitrary invertible matrix the above formula is then proved by first diagonalising the matrix.

Variation of this formula yields

$$\frac{1}{\det M} \delta(\det M) = \mathrm{Tr}\ \left(M^{-1} \delta M \right).$$

Let us now specialise and take the metric $g_{\mu\nu}$ for the matrix M. This yields

$$\delta g = g \left(g^{\mu\nu} \delta g_{\mu\nu} \right) = -g \left(g_{\mu\nu} \delta g^{\mu\nu} \right).$$

Hence

$$\delta\sqrt{-g} = -\frac{1}{2\sqrt{-g}} \delta g = \frac{1}{2} \frac{g}{\sqrt{-g}} g_{\mu\nu} \delta g^{\mu\nu} = -\frac{1}{2} \sqrt{-g} g_{\mu\nu} \delta g^{\mu\nu}.$$

Putting everything together, we obtain the variation of the Einstein-Hilbert action

$$\delta S_{EH} = -\frac{c^3}{16\pi G} \int d^4x \sqrt{-g} \left[Ric_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R - \Lambda g_{\mu\nu} \right] \delta g^{\mu\nu}.$$

Requiring that the variation of the action vanishes for arbitrary variations $\delta g^{\mu\nu}$ implies

$$Ric_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R - \Lambda g_{\mu\nu} = 0.$$

These are Einstein's equations in the case that no additional matter fields are present.

6.6 The energy-momentum tensor of general relativity

In the presence of additional fields and matter, the total action is given by

$$S = S_{EH} + S_{\text{particle}} + S_{\text{fields}} + \ldots$$

with

$$\begin{aligned} S_{\text{particle}} &= -mc \int\limits_a^b ds, \\ S_{\text{fields}} &= -\frac{1}{16\pi c} \int d^4x \sqrt{-g} F_{\mu\nu} F^{\mu\nu}, . \end{aligned}$$

Einstein's equations contain the energy-momentum tensor. In our review of classical field theory we have already seen a general method to compute the energy-momentum tensor from a Lagrange density $\mathscr{L}(\phi, \partial_\mu \phi)$:

$$T^{\mu\nu} = \left(\frac{\partial \mathscr{L}}{\partial \left(\partial_\mu \phi\right)} \partial^\nu \phi(x) \right) - g^{\mu\nu} \mathscr{L} + \partial_\rho B^{\mu\rho\nu}.$$

In this formula $B^{\mu\rho\nu}$ is anti-symmetric in μ and ρ and determined such that $T^{\mu\nu}$ is symmetric. Example: Consider a scalar field with Lagrange density

$$\mathscr{L} = \frac{\hbar^2 c}{2} \left[g^{\mu\nu} \left(\partial_\mu \phi(x)\right) \left(\partial_\nu \phi(x)\right) - \frac{m^2 c^2}{\hbar^2} \left(\phi(x)\right)^2 \right].$$

One finds ($\partial_\rho B^{\mu\rho\nu}$ is vanishing in this case):

$$\begin{aligned} T^{\mu\nu} &= \left(\frac{\partial \mathscr{L}}{\partial \left(\partial_\mu \phi\right)} \partial^\nu \phi(x) \right) - g^{\mu\nu} \mathscr{L} \\ &= \frac{\hbar^2 c}{2} \left[2 \left(\partial^\mu \phi(x)\right) \left(\partial^\nu \phi(x)\right) - g^{\mu\nu} \left(\partial_\lambda \phi(x)\right) \left(\partial^\lambda \phi(x)\right) + \frac{m^2 c^2}{\hbar^2} g^{\mu\nu} \left(\phi(x)\right)^2 \right]. \end{aligned}$$

Let us now consider an alternative method to compute the energy-momentum tensor. This method has the advantage, that it gives directly the correct and symmetric result. We consider the action

$$S = \frac{1}{c} \int d^4x \sqrt{-g} \mathscr{L}.$$

Variation with respect to $g^{\mu\nu}$ yields

$$\begin{aligned} \delta S &= \frac{1}{c} \int d^4x \left[\frac{\partial \sqrt{-g} \mathscr{L}}{\partial g^{\mu\nu}} \delta g^{\mu\nu} + \frac{\partial \sqrt{-g} \mathscr{L}}{\partial \frac{\partial g^{\mu\nu}}{\partial x^\lambda}} \delta \frac{g^{\mu\nu}}{\partial x^\lambda} \right] \\ &= \frac{1}{c} \int d^4x \left[\frac{\partial \sqrt{-g} \mathscr{L}}{\partial g^{\mu\nu}} - \frac{\partial}{\partial x^\lambda} \frac{\partial \sqrt{-g} \mathscr{L}}{\partial \frac{\partial g^{\mu\nu}}{\partial x^\lambda}} \right] \delta g^{\mu\nu}. \end{aligned}$$

We set

$$\frac{1}{2}\sqrt{-g}T_{\mu\nu} = \frac{\partial\sqrt{-g}\mathscr{L}}{\partial g^{\mu\nu}} - \frac{\partial}{\partial x^\lambda}\frac{\partial\sqrt{-g}\mathscr{L}}{\partial\frac{\partial g^{\mu\nu}}{\partial x^\lambda}}.$$

This yields

$$\delta S = \frac{1}{2c}\int d^4x\,\sqrt{-g}T_{\mu\nu}\delta g^{\mu\nu}.$$

It can be shown that

$$T_{\mu\nu} = \frac{2}{\sqrt{-g}}\left[\frac{\partial\sqrt{-g}\mathscr{L}}{\partial g^{\mu\nu}} - \frac{\partial}{\partial x^\lambda}\frac{\partial\sqrt{-g}\mathscr{L}}{\partial\frac{\partial g^{\mu\nu}}{\partial x^\lambda}}\right]$$

agrees with the first definition of the energy-momentum tensor. Let us verify this for the example of a scalar field discussed above:

$$\mathscr{L} = \frac{\hbar^2 c}{2}\left[g^{\mu\nu}\left(\partial_\mu\phi(x)\right)\left(\partial_\nu\phi(x)\right) - \frac{m^2c^2}{\hbar^2}\left(\phi(x)\right)^2\right].$$

We find

$$\begin{aligned}
T_{\mu\nu} &= \frac{2}{\sqrt{-g}}\frac{\partial\sqrt{-g}\mathscr{L}}{\partial g^{\mu\nu}} = 2\frac{\partial\mathscr{L}}{\partial g^{\mu\nu}} + \frac{2}{\sqrt{-g}}\mathscr{L}\frac{\partial\sqrt{-g}}{\partial g^{\mu\nu}} \\
&= 2\frac{\partial\mathscr{L}}{\partial g^{\mu\nu}} - \mathscr{L}g_{\mu\nu} \\
&= \frac{\hbar^2 c}{2}\left[2\left(\partial_\mu\phi(x)\right)\left(\partial_\nu\phi(x)\right) - g_{\mu\nu}\left(\partial_\lambda\phi(x)\right)\left(\partial^\lambda\phi(x)\right) + \frac{m^2c^2}{\hbar^2}g_{\mu\nu}\left(\phi(x)\right)^2\right].
\end{aligned}$$

Let us return to the general case. We obtain for the variation of

$$S = -\frac{c^3}{16\pi G}\int d^4x\,\sqrt{-g}\,(R+2\Lambda) + \frac{1}{c}\int d^4x\,\sqrt{-g}\mathscr{L}$$

the expression

$$\delta S = -\frac{c^3}{16\pi G}\int d^4x\,\sqrt{-g}\left[Ric_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R - \Lambda g_{\mu\nu}\right]\delta g^{\mu\nu} + \frac{1}{2c}\int d^4x\,\sqrt{-g}T_{\mu\nu}\delta g^{\mu\nu}.$$

Hence

$$-\frac{c^3}{16\pi G}\left[Ric_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R - \Lambda g_{\mu\nu}\right] + \frac{1}{2c}T_{\mu\nu} = 0,$$

or

$$Ric_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R - \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4}T_{\mu\nu}.$$

Let us now discuss **conservation laws associated to the energy-momentum tensor**. The energy-momentum tensor satisfies

$$\nabla_\mu T^{\mu\nu} \;=\; 0.$$

In our heuristic motivation for Einstein's equations we used this as an input. However, if we accept the Einstein-Hilbert action as a starting point, this equation follows from Einstein's equations and $\nabla_\mu G^{\mu\nu} = 0$, the latter is due to the Bianchi identity. Let us first consider a vector j^μ, which satisfies

$$\nabla_\mu j^\mu \;=\; 0$$

and vanishes at spatial infinity. Stoke's theorem reads

$$\int\limits_M d^4x\sqrt{|g|}\nabla_\mu V^\mu \;=\; \int\limits_{\partial M} d^3y\sqrt{|\gamma|}\, n_\mu V^\mu,$$

and setting $V^\mu = j^\mu$ yields

$$0 \;=\; \int\limits_{\partial M} d^3y\sqrt{|\gamma|}\, n_\mu j^\mu.$$

Taking M as the region bounded by the time coordinates t_i and t_f and extending to spatial infinity one finds the conservation law

$$\int\limits_{t=t_f} d^3y\sqrt{|\gamma|}\, j^0 - \int\limits_{t=t_i} d^3y\sqrt{|\gamma|}\, j^0 \;=\; 0.$$

Note that Stoke's theorem requires a vector V^μ, we cannot plug in a tensor $T^{\mu\nu}$ or some fixed components of a tensor like $T^{\mu 0}$. However, the contraction of the rank 2 tensor $T^{\mu\nu}$ with a vector ξ_ν transforms as a vector. Let us now investigate under which conditions we have $\nabla_\mu(T^{\mu\nu}\xi_\nu) = 0$:

$$\begin{aligned}
\nabla_\mu\left(T^{\mu\nu}\xi_\nu\right) &= \frac{1}{2}\nabla_\mu\left(T^{\mu\nu}\xi_\nu\right) + \frac{1}{2}\nabla_\nu\left(T^{\mu\nu}\xi_\mu\right) \\
&= \frac{1}{2}\left(\nabla_\mu T^{\mu\nu}\right)\xi_\nu + \frac{1}{2}T^{\mu\nu}\nabla_\mu\xi_\nu + \frac{1}{2}\left(\nabla_\nu T^{\mu\nu}\right)\xi_\mu + \frac{1}{2}T^{\mu\nu}\nabla_\nu\xi_\mu \\
&= \frac{1}{2}T^{\mu\nu}\left(\nabla_\mu\xi_\nu + \nabla_\nu\xi_\mu\right).
\end{aligned}$$

Thus $\nabla_\mu(T^{\mu\nu}\xi_\nu) = 0$ if $\nabla_\mu\xi_\nu + \nabla_\nu\xi_\mu = 0$ or in other words if ξ_ν is a Killing vector field. If we now assume that ξ_ν is a Killing vector field and $T^{\mu\nu}\xi_\nu$ vanishes at spatial infinity we obtain with the same reasoning as above the conservation law

$$\int\limits_{t=t_f} d^3y\sqrt{|\gamma|}\, T^{0\nu}\xi_\nu - \int\limits_{t=t_i} d^3y\sqrt{|\gamma|}\, T^{0\nu}\xi_\nu \;=\; 0.$$

If $\xi_\nu = (1,\vec{0})$ is a Killing vector field, we have energy conservation in the usual form

$$\int\limits_{t=t_f} d^3y \sqrt{|\gamma|}\, T^{00} - \int\limits_{t=t_i} d^3y \sqrt{|\gamma|}\, T^{00} \;=\; 0.$$

saying that the integral over the energy density over spatial space is conserved. Note that $\nabla_\mu T^{\mu\nu}$ alone is not enough to obtain this result, we need in addition that $\xi_\nu = (1,\vec{0})$ is a Killing vector field. This is of course in accordance with Noether's theorem: A Killing vector field generates a symmetry (in this case time translation) and only if the system is invariant under time translations energy conservation follows.

6.7 The Palatini formalism

Preliminary remark: Let us consider within classical mechanics the action

$$S \;=\; \int_{t_a}^{t_b} L(q,\dot{q})dt, \qquad L(q,\dot{q}) = \frac{1}{2}\dot{q}^2 - V(q).$$

Variation with respect to the generalised coordinate $q(t)$ and keeping the end-points fixed $\delta q(t_a) = \delta q(t_b) = 0$ yields the Euler-Lagrange equation

$$\frac{\delta L}{\delta q} - \frac{d}{dt}\frac{\delta L}{\delta \dot{q}} \;=\; 0, \qquad \ddot{q} = -\frac{\delta V}{\delta q}.$$

This is the formulation of classical mechanics according to Lagrange. Equally well we may consider the Hamiltonian formulation of classical mechanics:

$$S \;=\; \int_{t_a}^{t_b} \left(p\dot{q} - H(q,p)\right) dt, \qquad H(q,p) = \frac{1}{2}p^2 + V(q)$$

We now consider $q(t)$ and $p(t)$ as independent (i.e. we do not set from the beginning $p(t) = \dot{q}(t)$) and vary with respect to $q(t)$ and $p(t)$. Variation with respect to $p(t)$ yields the relation

$$\dot{q} \;=\; \frac{\delta H(q,p)}{\delta p} = p.$$

Variation with respect to $q(t)$ yields the equation of motion

$$\dot{p} \;=\; -\frac{\delta H(q,p)}{\delta q} = -\frac{\delta V}{\delta q}.$$

Let us now transfer this to general relativity. For the derivation of Einstein's equations from the Einstein-Hilbert action

$$S_{EH} \;=\; -\frac{c^3}{16\pi G}\int d^4x\, \sqrt{-g}\left(g^{\mu\nu} Ric_{\mu\nu} + 2\Lambda\right)$$

we considered the variation with respect to the inverse metric $g^{\mu\nu}$. The Ricci tensor

$$Ric_{\mu\nu} = \partial_\kappa \Gamma^\kappa_{\nu\mu} - \partial_\nu \Gamma^\kappa_{\kappa\mu} + \Gamma^\eta_{\nu\mu}\Gamma^\kappa_{\kappa\eta} - \Gamma^\eta_{\kappa\mu}\Gamma^\kappa_{\nu\eta}$$

depends on the Christoffel symbols, which in turn depend on the metric

$$\Gamma^\kappa_{\mu\nu} = \frac{1}{2}g^{\kappa\lambda}\left(\partial_\mu g_{\nu\lambda} + \partial_\nu g_{\mu\lambda} - \partial_\lambda g_{\mu\nu}\right).$$

Within the Palatini formalism we consider the (inverse) metric $g^{\mu\nu}$ and the (symmetric) connection coefficients as **independent** quantities. Variation with respect to the inverse metric yields Einstein's equations

$$Ric_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R - \Lambda g_{\mu\nu} = 0.$$

(Here only the terms $(\delta S)_2$ and $(\delta S)_3$ contribute, which give the variation of $g^{\mu\nu}$ and $\sqrt{-g}$ with respect to the inverse metric $g^{\mu\nu}$.) Within the Palatini formalism the Ricci tensor $Ric_{\mu\nu}$ depends only on the connection coefficients. The variation of

$$Ric_{\mu\nu} = \partial_\kappa C^\kappa_{\nu\mu} - \partial_\nu C^\kappa_{\kappa\mu} + C^\eta_{\nu\mu}C^\kappa_{\kappa\eta} - C^\eta_{\kappa\mu}C^\kappa_{\nu\eta}$$

with respect to the connection coefficients $C^\kappa_{\mu\nu}$ yields

$$\delta Ric_{\mu\nu} = \nabla_\kappa \delta C^\kappa_{\nu\mu} - \nabla_\nu \delta C^\kappa_{\kappa\mu}.$$

Therefore we obtain for the variation of the action with respect to the connection coefficients:

$$\begin{aligned}
\delta S_{EH} &= -\frac{c^3}{16\pi G}\int d^4x\,\sqrt{-g}g^{\mu\nu}\delta Ric_{\mu\nu} = -\frac{c^3}{16\pi G}\int d^4x\,\sqrt{-g}g^{\mu\nu}\left(\nabla_\kappa\delta C^\kappa_{\nu\mu} - \nabla_\nu\delta C^\kappa_{\kappa\mu}\right) \\
&= \frac{c^3}{16\pi G}\int d^4x\,\left(\nabla_\kappa\sqrt{-g}g^{\mu\nu} - \delta^\nu_\kappa\nabla_\lambda\sqrt{-g}g^{\mu\lambda}\right)\delta C^\kappa_{\nu\mu}.
\end{aligned}$$

This has to hold for arbitrary variations, hence the expression in the bracket has to vanish. This implies that the symmetric combination has to vanish as well:

$$\nabla_\kappa\sqrt{-g}g^{\mu\nu} - \frac{1}{2}\delta^\nu_\kappa\nabla_\lambda\sqrt{-g}g^{\mu\lambda} - \frac{1}{2}\delta^\mu_\kappa\nabla_\lambda\sqrt{-g}g^{\nu\lambda} = 0.$$

This is a system of 40 equations for the 40 covariant derivatives $\nabla_\kappa\sqrt{-g}g^{\mu\nu}$. The unique solution is

$$\nabla_\kappa\sqrt{-g}g^{\mu\nu} = 0.$$

One then shows

$$\nabla_\kappa\sqrt{-g} = 0,$$

this implies immediately

$$\nabla_\kappa g^{\mu\nu} = 0.$$

With the help of

$$0 = \nabla_\kappa \sqrt{-g}\delta^\rho_\nu = \nabla_\kappa \sqrt{-g} g^{\rho\mu} g_{\mu\nu} = \sqrt{-g} g^{\rho\mu} \nabla_\kappa g_{\mu\nu}$$

it follows that

$$\nabla_\kappa g_{\mu\nu} = 0.$$

We recognise this equation as the condition that the metric is covariantly constant with respect to the connection. Together with the assumption that the connection is torsion free (symmetric), this uniquely defines the Levi-Civita connection. In this case the connection coefficients are given by the Christoffel symbols.

Remark: In the case where one considers only the metric as an independent field, the Einstein-Hilbert action contains second derivatives of the metric. The advantage of the Palatini formalism is given by the fact, that the action contains in this formalism only first derivatives of the connection coefficients.

6.8 The vielbein formalism

The vielbein formalism is required to describe the interaction of fermions with gravitation.

We start with a manifold of dimension n. Up to now we used as basis vectors for the tangent space at the point p the derivatives in the direction of the coordinate axes:

$$e_\mu = \partial_\mu.$$

As standard basis for the cotangent space at the point p we used up to now the corresponding dual vectors:

$$\theta^\mu = dx^\mu.$$

Let us look at an example: The (two-dimensional) surface of a sphere with coordinates given by a polar angle ϑ and an azimuthal angle φ. The metric in these coordinates reads

$$g = d\vartheta \otimes d\vartheta + \sin^2 \vartheta d\varphi \otimes d\varphi.$$

At the point $(\vartheta, \varphi) = (\pi/3, 0)$ we find

$$g(e_\varphi, e_\varphi) = \frac{3}{4},$$

whereas at the point $(\vartheta,\varphi) = (\pi/2,0)$ we obtain

$$g(e_\varphi,e_\varphi) \;=\; 1.$$

More generally there can be the case, that two basis vectors are orthogonal at point A, but not at point B. This happens for example if we consider a metric containing a term $c(x)e_i \otimes e_j$, where the coefficient $c(x)$ is vanishing at point A, but not at point B. We see that the derivatives in the direction of the coordinate axes generally do not form an orthonormal basis. For the tangent space we may define a new basis e_a, which by definition satisfies

$$g(e_a,e_b) \;=\; \eta_{ab}.$$

(This is the appropriate definition for a Lorentzian manifold, for a manifold with Euclidean signature one replaces η_{ab} by δ_{ab}.) In general, this basis is no longer given by the derivatives in the direction of the coordinate axes, but we may express the new basis as a linear combination of the old basis e_μ:

$$e_a \;=\; e_a^{\;\mu}\, e_\mu,$$

where $e_a^{\;\mu}$ is an invertible $n\times n$-matrix. In order to preserve the orientation we require in addition $\det e_a^{\;\mu} > 0$. The new basis e_a is called the **non-coordinate basis**. A widely adopted convention uses greek indices for the coordinate basis e_μ and latin indices for the non-coordinate basis e_a. Furthermore, one sometimes refers to e_μ as a **holonomic basis**, and to e_a as an **anholonomic basis**. The $n\times n$-matrix $e_a^{\;\mu}$ is called generally the **vielbein**, on a manifold of dimension four the vierbein (and on a manifold of dimension three the dreibein etc.). We denote by $e^a_{\;\mu}$ the inverse matrix of $e_a^{\;\mu}$:

$$e_a^{\;\mu} e^a_{\;\nu} = \delta^\mu_\nu, \qquad e_a^{\;\mu} e^b_{\;\mu} = \delta^b_a.$$

With the help of $e^a_{\;\mu}$ we obtain

$$e_\mu \;=\; e^a_{\;\mu} e_a$$

and

$$g_{\mu\nu} \;=\; e^a_{\;\mu} e^b_{\;\nu} \eta_{ab}.$$

In addition we may define a new basis θ^a for the cotangent space as the dual basis to the non-coordinate basis e_a:

$$\langle \theta^a, e_b\rangle \;=\; \delta^a_b.$$

One finds

$$\theta^a = e^a_{\;\mu}\theta^\mu, \qquad \theta^\mu = e_a^{\;\mu}\theta^a.$$

Previously we introduced the Lie bracket for vector fields, which yields again a vector field:

$$[X,Y] = \left(X^\mu \partial_\mu Y^\nu - Y^\mu \partial_\mu X^\nu\right) e_\nu$$

For the coordinate basis we have

$$\left[e_\mu, e_\nu\right] = 0.$$

However, for the non-coordinate basis we obtain

$$[e_a, e_b] = \left[e_a{}^\mu e_\mu, e_b{}^\nu e_\nu\right] = \left(e_a{}^\mu \partial_\mu e_b{}^\nu - e_b{}^\mu \partial_\mu e_a{}^\nu\right) e_\nu = c_{ab}{}^c e_c,$$

with

$$c_{ab}{}^c = \left(e_a{}^\mu \partial_\mu e_b{}^\nu - e_b{}^\mu \partial_\mu e_a{}^\nu\right) e^c{}_\nu,$$

i.e. the non-coordinate basis has a non-vanishing Lie bracket:

$$[e_a, e_b] = c_{ab}{}^c e_c.$$

With the help of $e^a{}_\mu$ and $e_a{}^\mu$ we may convert tensors from the coordinate basis to the non-coordinate basis and vice versa. For example, a $(1,2)$-tensor in the non-coordinate basis is converted to the coordinate basis by

$$T^\kappa_{\mu\nu} = e_c{}^\kappa e^a{}_\mu e^b{}_\nu T^c_{ab}.$$

The connection coefficients do not form a tensor and we write

$$\nabla_a e_b = \omega^c_{ab} e_c.$$

We have

$$\begin{aligned}
\nabla_a e_b &= \nabla_{e_a{}^\mu e_\mu}\left(e_b{}^\nu e_\nu\right) = e_a{}^\mu \nabla_\mu \left(e_b{}^\nu e_\nu\right) = e_a{}^\mu \left[\left(\partial_\mu e_b{}^\nu\right) e_\nu + C^\kappa_{\mu\nu} e_b{}^\nu e_\kappa\right] \\
&= e_a{}^\mu \left[\left(\partial_\mu e_b{}^\nu\right) e^c{}_\nu + C^\kappa_{\mu\nu} e_b{}^\nu e^c{}_\kappa\right] e_c = e_a{}^\mu e^c{}_\nu \left[\partial_\mu e_b{}^\nu + C^\nu_{\mu\rho} e_b{}^\rho\right] e_c,
\end{aligned}$$

and therefore

$$\omega^c_{ab} = e_a{}^\mu e^c{}_\nu \left[\partial_\mu e_b{}^\nu + C^\nu_{\mu\rho} e_b{}^\rho\right].$$

We define the connection one-form $\omega^a{}_b$ by

$$\omega^a{}_b = \omega^a_{cb} \theta^c = e^a{}_\nu \left(\partial_\mu e_b{}^\nu + C^\nu{}_{\mu\rho} e_b{}^\rho\right) dx^\mu.$$

The one-form $\omega^a{}_b$ is also known as the **spin connection** one-form.

Let us now consider the torsion tensor and the curvature tensor in the non-coordinate basis:

$$\begin{aligned}
T(e_a, e_b) &= T^c_{ab} e_c, \\
R(e_a, e_b, e_c) &= R^d{}_{cab} e_d.
\end{aligned}$$

We determine T^c_{ab} from the definition of the torsion tensor:

$$T(e_a, e_b) = \nabla_a e_b - \nabla_b e_a - [e_a, e_b] = (\omega^c{}_{ab} - \omega^c{}_{ba} - c_{ab}{}^c)\, e_c,$$

hence

$$T^c_{ab} = \omega^c{}_{ab} - \omega^c{}_{ba} - c_{ab}{}^c.$$

In the same way we obtain from

$$\begin{aligned} R(e_a, e_b, e_c) &= \nabla_a \nabla_b e_c - \nabla_b \nabla_a - \nabla_{[e_a, a_b]} e_c \\ &= \left(\partial_a \omega^d{}_{bc} - \partial_b \omega^d{}_{ac} + \omega^e{}_{bc}\omega^d{}_{ae} - \omega^e{}_{ac}\omega^d{}_{be} - c_{ab}{}^e \omega^d{}_{ec}\right) e_d \end{aligned}$$

the coefficients

$$R^d_{cab} = \partial_a \omega^d{}_{bc} - \partial_b \omega^d{}_{ac} + \omega^e{}_{bc}\omega^d{}_{ae} - \omega^e{}_{ac}\omega^d{}_{be} - c_{ab}{}^e \omega^d{}_{ec}.$$

This allows us to define a torsion two-form T^a and a curvature two-form R^a_b:

$$\begin{aligned} T^a &= \frac{1}{2} T^a_{bc} \theta^b \wedge \theta^c, \\ R^a_b &= \frac{1}{2} R^a_{bcd} \theta^c \wedge \theta^d. \end{aligned}$$

With the help of these definitions we may now state the **structure equations of Cartan**:

$$\begin{aligned} T^a &= d\theta^a + \omega^a{}_b \wedge \theta^b, \\ R^a_b &= d\omega^a{}_b + \omega^a{}_c \wedge \omega^c{}_b. \end{aligned}$$

Let us also consider the Bianchi identities in the non-coordinate basis:

$$\begin{aligned} dT^a + \omega^a{}_b \wedge T^b &= R^a_b \wedge e^b, \\ dR^a_b + \omega^a{}_c \wedge R^c_b - R^a_c \wedge \omega^c{}_b &= 0. \end{aligned}$$

Remark: Previously we proved the Bianchi identities in the coordinate basis for the case that the torsion tensor is vanishing. The form of the Bianchi identities stated above holds in general (and in particular also for $T^a \neq 0$).

The vielbein formalism allows for an elegant formulation of general relativity. In addition, the vielbein formalism has the advantage that spinor fields can be included. Instead of the metric $g_{\mu\nu}$ one uses within the vielbein formalism the vielbein $e_a{}^\mu$ and the spin connection $\omega^c{}_{ab}$ as fundamental fields. Similar to the Palatini formalism (which uses the inverse metric and the symmetric connection coefficients as fundamental fields) one may show that within the vielbein formalism the spin connection one-form may be expressed in terms of the vielbein fields. Within the vielbein formalism and within the Palatini formalism we obtain instead of second order differential equations a system of coupled first order differential equations.

The starting point for the formulation of general relativity within the vielbein formalism are two one-forms. We consider the vielbein field

$$\theta^a = e^a{}_\mu dx^\mu$$

and the spin connection one-form

$$\omega^a{}_b = \omega^a_{\mu b} dx^\mu.$$

We require that the transformation from the coordinate basis to the non-coordinate basis is invertible and orientation-preserving. This translates to the requirement

$$\det \left(e^a{}_\mu\right) > 0.$$

The spin connection defines the covariant derivative:

$$\nabla_\mu e_a = \omega^b{}_{\mu a} e_b.$$

Torsion and curvature are given by

$$\begin{aligned} T^a &= d\theta^a + \omega^a{}_b \wedge \theta^b, \\ R^a_b &= d\omega^a{}_b + \omega^a{}_c \wedge \omega^c{}_b. \end{aligned}$$

Explicitly, we find for the curvature

$$\begin{aligned} R^a_b &= \frac{1}{2} R^a{}_{b\mu\nu} dx^\mu \wedge dx^\nu, \\ R^a{}_{b\mu\nu} &= \partial_\mu \omega^a{}_{\nu b} - \partial_\nu \omega^a{}_{\mu b} + \omega^a{}_{\mu c} \omega^c{}_{\nu b} - \omega^a{}_{\nu c} \omega^c{}_{\mu b}. \end{aligned}$$

The metric is given by

$$g_{\mu\nu} = e^a{}_\mu e^b{}_\nu \eta_{ab}.$$

The vielbein defines a unique torsion-free and metric-compatible spin connection. This is most easily seen as follows: The relation between the connection coefficients $\omega^a_{\mu b}$ in the non-coordinate basis and the connection coefficients $C^\kappa_{\mu\nu}$ in the coordinate basis is given by

$$\omega^a_{\mu b} = e^a{}_\nu \left[\partial_\mu e_b{}^\nu + C^\nu_{\mu\rho} e_b{}^\rho\right].$$

The connection in the coordinate basis should be torsion-free and metric-compatible, hence it must be the Levi-Civita connection. The Levi-Civita connection is given in terms of derivatives of the metric as

$$\begin{aligned} C^\kappa{}_{\mu\nu} &= \Gamma^\kappa{}_{\mu\nu} = \frac{1}{2} g^{\kappa\lambda} \left(\partial_\mu g_{\nu\lambda} + \partial_\nu g_{\mu\lambda} - \partial_\lambda g_{\mu\nu}\right) \\ &= \frac{1}{2} e_a{}^\kappa e^{a\lambda} \left[e_{b\lambda}\left(\partial_\mu e^b{}_\nu + \partial_\nu e^b{}_\mu\right) + e_{b\nu}\left(\partial_\mu e^b{}_\lambda - \partial_\lambda e^b{}_\mu\right) + e_{b\mu}\left(\partial_\nu e^b{}_\lambda - \partial_\lambda e^b{}_\nu\right)\right]. \end{aligned}$$

We therefore obtain

$$\omega^a_{\mu b} = \frac{1}{2} e_b{}^\nu e^{a\lambda} \left[e_{c\mu} \left(\partial_\nu e^c{}_\lambda - \partial_\lambda e^c{}_\nu \right) + e_{c\nu} \left(\partial_\mu e^c{}_\lambda - \partial_\lambda e^c{}_\mu \right) - e_{c\lambda} \left(\partial_\mu e^c{}_\nu - \partial_\nu e^c{}_\mu \right) \right].$$

For the action we find

$$\begin{aligned} S_{EH} &= -\frac{c^3}{16\pi G} \int d^4x \sqrt{-g} \left(R + 2\Lambda \right) \\ &= -\frac{c^3}{16\pi G} \int \varepsilon_{abcd} \left(\frac{1}{2} \theta^a \wedge \theta^b \wedge R^{cd} + \frac{\Lambda}{12} \theta^a \wedge \theta^b \wedge \theta^c \wedge \theta^d \right). \end{aligned}$$

For the derivation of the last line let us first consider the term with the cosmological constant. Here we used

$$\begin{aligned} \frac{\Lambda}{12} \varepsilon_{abcd} \theta^a \wedge \theta^b \wedge \theta^c \wedge \theta^d &= \frac{\Lambda}{12} \varepsilon_{abcd} e^a{}_\mu e^b{}_\nu e^c{}_\rho e^d{}_\sigma dx^\mu \wedge dx^\nu \wedge dx^\rho \wedge dx^\sigma \\ &= -\frac{\Lambda}{12} \varepsilon_{abcd} \varepsilon^{\mu\nu\rho\sigma} e^a{}_\mu e^b{}_\nu e^c{}_\rho e^d{}_\sigma dx^0 \wedge dx^1 \wedge dx^2 \wedge dx^3 \\ &= 2\Lambda \det \left(e^a{}_\mu \right) \, dx^0 \wedge dx^1 \wedge dx^2 \wedge dx^3, \end{aligned}$$

with

$$dx^\mu \wedge dx^\nu \wedge dx^\rho \wedge dx^\sigma = -\varepsilon^{\mu\nu\rho\sigma} dx^0 \wedge dx^1 \wedge dx^2 \wedge dx^3.$$

The minus sign is due to our convention $\varepsilon_{0123} = 1$ which implies $\varepsilon^{0123} = -1$ On the other hand we also have

$$\sqrt{-g} = \sqrt{-\det g_{\mu\nu}} = \sqrt{-\det \left(e^a{}_\mu e^b{}_\nu \eta_{ab} \right)} = \sqrt{-\det e^a{}_\mu \det e^b{}_\nu \det \eta_{ab}} = \det \left(e^a{}_\mu \right),$$

which shows the equality of the two terms proportional to the cosmological constant.
In order to derive the term involving the curvature form we need the Schouten identity:

$$\varepsilon_{abcd} e_f^{\ \mu} + \varepsilon_{bcdf} e_a^{\ \mu} + \varepsilon_{cdfa} e_b^{\ \mu} + \varepsilon_{dfab} e_c^{\ \mu} + \varepsilon_{fabc} e_d^{\ \mu} = 0.$$

Let us now take the action

$$S_{EH} = -\frac{c^3}{16\pi G} \int \varepsilon_{abcd} \left(\frac{1}{2} \theta^a \wedge \theta^b \wedge R^{cd} + \frac{\Lambda}{12} \theta^a \wedge \theta^b \wedge \theta^c \wedge \theta^d \right)$$

together with the constraints

$$\det \left(e^a{}_\mu \right) > 0, \qquad \omega^a_{\ \mu}{}^b = -\omega^b_{\ \mu}{}^a$$

as starting point. The anti-symmetry of the spin connection $\omega^a_{\ \mu}{}^b = -\omega^b_{\ \mu}{}^a$ implies $\nabla_\kappa g_{\mu\nu} = 0$. This is easily shown as follows:

$$\begin{aligned} 0 &= \nabla_\mu \left(g_{\rho\sigma} dx^\rho \otimes dx^\sigma \right) = \nabla_\mu \left(\eta_{ab} \theta^a \otimes \theta^b \right) = -\left(\omega_{a\mu b} + \omega_{b\mu a} \right) \theta^a \otimes \theta^b \\ &= -\left(\omega^a_{\ \mu}{}^b + \omega^b_{\ \mu}{}^a \right) \eta_{ac} \eta_{bd} \theta^c \otimes \eta^d. \end{aligned}$$

Variation of the action with respect to the spin connection yields

$$\begin{aligned}\delta S_{EH} &= -\frac{c^3}{16\pi G}\int dx^\mu\wedge dx^\nu\wedge dx^\rho\wedge dx^\sigma \varepsilon_{cdef}\eta^{fg}\eta_{bh}\left[\frac{1}{2}\left(\partial_\nu e^c{}_\rho\right)e^d{}_\sigma\delta^e_a\delta^h_g\right.\\ &\left. +\frac{1}{2}e^c{}_\rho\left(\partial_\nu e^d{}_\sigma\right)\delta^e_a\delta^h_g-\frac{1}{2}e^c{}_\rho e^d{}_\sigma\omega^e{}_{\nu a}\delta^h_g+\frac{1}{2}e^c{}_\rho e^d{}_\sigma\omega^h{}_{\nu g}\delta^e_a\right]\delta\omega^{a\ b}_{\ \mu}.\end{aligned}$$

Using the anti-symmetry of the spin connection $\omega^{a\ b}_{\ \mu}=-\omega^{b\ a}_{\ \mu}$ this implies that the following expression, anti-symmetric in a and b, has to vanish:

$$\begin{aligned}0 &= -\frac{1}{4}\varepsilon^{\mu\nu\rho\sigma}\varepsilon_{cdef}\left\{\left(\delta^e_a\delta^f_b-\delta^e_b\delta^f_a\right)\left[\left(\partial_\nu e^c{}_\rho\right)e^d{}_\sigma+e^c{}_\rho\left(\partial_\nu e^d{}_\sigma\right)\right]\right.\\ &\quad\left.-e^c{}_\rho e^d{}_\sigma\left[\omega^e{}_{\nu a}\delta^f_b+\omega^f{}_{\nu b}\delta^e_a-\omega^e{}_{\nu b}\delta^f_a-\omega^f{}_{\nu a}\delta^e_b\right]\right\}\\ &= -\frac{1}{2}\varepsilon^{\mu\nu\rho\sigma}\varepsilon_{cdef}\left\{\delta^e_a\delta^f_b\left[\left(\partial_\nu e^c{}_\rho\right)e^d{}_\sigma+e^c{}_\rho\left(\partial_\nu e^d{}_\sigma\right)\right]-e^c{}_\rho e^d{}_\sigma\left[\omega^e{}_{\nu a}\delta^f_b-\omega^e{}_{\nu b}\delta^f_a\right]\right\}\\ &= -\frac{1}{2}\varepsilon^{\mu\nu\rho\sigma}\varepsilon_{cdef}\left\{\delta^e_a\delta^f_b\left(\partial_\nu e^c{}_\rho-\partial_\rho e^c{}_\nu\right)-e^e{}_\rho\left(\delta^f_a\delta^g_b-\delta^g_a\delta^f_b\right)\omega^c{}_{\nu g}\right\}e^d{}_\sigma.\end{aligned}$$

We have

$$\delta^f_a\delta^g_b-\delta^g_a\delta^f_b = -\frac{1}{2}\varepsilon_{abij}\varepsilon^{fgij}$$

and

$$\begin{aligned}\varepsilon_{cdef}\left(\delta^f_a\delta^g_b-\delta^g_a\delta^f_b\right) &= \frac{1}{2}\varepsilon_{abij}\varepsilon^{gijf}\varepsilon_{cdef}\\ &= -\frac{1}{2}\varepsilon_{abij}\left(\delta^g_c\delta^i_d\delta^j_e+\delta^i_c\delta^j_d\delta^g_e+\delta^j_c\delta^g_d\delta^i_e-\delta^j_c\delta^i_d\delta^g_e-\delta^i_c\delta^g_d\delta^j_e-\delta^g_c\delta^j_d\delta^i_e\right)\\ &= -\varepsilon_{abij}\left(\delta^g_c\delta^i_d\delta^j_e+\delta^i_c\delta^j_d\delta^g_e+\delta^j_c\delta^g_d\delta^i_e\right).\end{aligned}$$

Therefore we obtain

$$\begin{aligned}0 &= -\frac{1}{2}\varepsilon^{\mu\nu\rho\sigma}\left\{\varepsilon_{abcd}\left(\partial_\nu e^c{}_\rho-\partial_\rho e^c{}_\nu\right)+e^e{}_\rho\varepsilon_{abij}\left(\delta^g_c\delta^i_d\delta^j_e+\delta^i_c\delta^j_d\delta^g_e+\delta^j_c\delta^g_d\delta^i_e\right)\omega^c{}_{\nu g}\right\}e^d{}_\sigma\\ &= -\frac{1}{2}\varepsilon^{\mu\nu\rho\sigma}\left\{\varepsilon_{abcd}\left(\partial_\nu e^c{}_\rho-\partial_\rho e^c{}_\nu\right)+\varepsilon_{abcd}\omega^c{}_{\nu g}e^g{}_\rho+e^e{}_\rho\varepsilon_{abec}\omega^c{}_{\nu d}\right\}e^d{}_\sigma\\ &= -\frac{1}{2}\varepsilon^{\mu\nu\rho\sigma}\varepsilon_{abcd}\left\{\partial_\nu e^c{}_\rho-\partial_\rho e^c{}_\nu+\omega^c{}_{\nu g}e^g{}_\rho-\omega^c{}_{\rho g}e^g{}_\nu\right\}e^d{}_\sigma\end{aligned}$$

This is nothing else than the condition that the torsion vanishes:

$$\partial_\nu e^c{}_\rho-\partial_\rho e^c{}_\nu+\omega^c{}_{\nu g}e^g{}_\rho-\omega^c{}_{\rho g}e^g{}_\nu = 0.$$

Variation of the action with respect to the vielbein field yields

$$\delta S_{EH} = -\frac{c^3}{16\pi G}\int dx^\mu\wedge dx^\nu\wedge dx^\rho\wedge dx^\sigma\varepsilon_{abcd}\left[\frac{1}{2}e^b{}_\nu R^{cd}{}_{\rho\sigma}+\frac{1}{3}\Lambda e^b{}_\nu e^c{}_\rho e^d{}_\sigma\right]\delta e^a{}_\mu.$$

This implies

$$0 = -\varepsilon^{\mu\rho\sigma\tau}\varepsilon_{abcd}\left[\frac{1}{2}e^b{}_\rho R^{cd}{}_{\sigma\tau} + \frac{1}{3}\Lambda e^b{}_\rho e^c{}_\sigma e^d{}_\tau\right].$$

Multiplication with $e^a{}_\kappa g^{\kappa\nu}$ yields

$$0 = -\varepsilon^{\mu\rho\sigma\tau}\varepsilon_{abcd}\left[\frac{1}{2}e^b{}_\rho R^{cd}{}_{\sigma\tau} + \frac{1}{3}\Lambda e^b{}_\rho e^c{}_\sigma e^d{}_\tau\right] e^a{}_\kappa g^{\kappa\nu}.$$

After a slightly lengthy calculation and by repeated use of the Schouten identity one finds

$$0 = -2\det\left(e^a{}_\rho\right)\left(Ric^{\mu\nu} - \frac{1}{2}g^{\mu\nu}R - \Lambda g^{\mu\nu}\right).$$

These are Einstein's field equations.

6.9 The Plebanski formalism

It is sometimes advantageous to work over the complex numbers instead of working over the real numbers. This is the main motivation for the Plebanski formalism. Within the Plebanski formalism we consider as within the vielbein formalism the vielbein and the spin connection as fundamental fields. In addition, we complexify the tangent space and the cotangent space. We then decompose all two-forms into a self dual and an anti-self dual part. We further postulate that gravity is only determined by the self dual part, i.e. we postulate that the anti-self dual part is vanishing.

Within the framework of the vielbein formalism we introduced at each point of space-time an anholonomic basis e_a of the tangent space. We are in particular interested in four-dimensional space-times. In this case the tangent space at a given point is a four-dimensional vector space. Within the Plebanski formalism we extend the vector space spanned by the vectors e_a from a real vector space to a complex vector space. In the same way we extend the cotangent space spanned by the cotangent basis vectors θ^a from a real vector space to a complex vector space.

The action of general relativity within the vielbein formalism is given by

$$S_{EH} = -\frac{c^3}{16\pi G}\int \varepsilon_{abcd}\left(\frac{1}{2}\theta^a\wedge\theta^b\wedge R^{cd} + \frac{\Lambda}{12}\theta^a\wedge\theta^b\wedge\theta^c\wedge\theta^d\right).$$

We may re-write this action in terms of two two-forms

$$\begin{aligned} B^{ab} &= \theta^a\wedge\theta^b = e^a{}_\mu e^b{}_\nu dx^\mu\wedge dx^\nu = \frac{1}{2}\left(e^a{}_\mu e^b{}_\nu - e^b{}_\mu e^a{}_\nu\right)dx^\mu\wedge dx^\nu, \\ R^{ab} &= \frac{1}{2}R^a{}_{c\mu\nu}\eta^{bc}dx^\mu\wedge dx^\nu. \end{aligned}$$

We obtain

$$S_{EH} = -\frac{c^3}{32\pi G}\int \varepsilon_{abcd}\left(B^{ab}\wedge R^{cd} + \frac{\Lambda}{6}B^{ab}\wedge B^{cd}\right).$$

We decompose the two-forms B^{ab} and R^{ab} into a self dual part and an anti-self dual part:

$$B^{ab} = B^{ab}_{\text{selfdual}} + B^{ab}_{\text{antiselfdual}}, \qquad R^{ab} = R^{ab}_{\text{selfdual}} + R^{ab}_{\text{antiselfdual}},$$

with

$$\begin{aligned} B^{ab}_{\text{selfdual}} &= \frac{1}{2}\left(B^{ab} + \frac{i}{2}\varepsilon^{ab}{}_{cd}B^{cd}\right), & B^{ab}_{\text{antiselfdual}} &= \frac{1}{2}\left(B^{ab} - \frac{i}{2}\varepsilon^{ab}{}_{cd}B^{cd}\right), \\ R^{ab}_{\text{selfdual}} &= \frac{1}{2}\left(R^{ab} + \frac{i}{2}\varepsilon^{ab}{}_{cd}R^{cd}\right), & R^{ab}_{\text{antiselfdual}} &= \frac{1}{2}\left(R^{ab} - \frac{i}{2}\varepsilon^{ab}{}_{cd}R^{cd}\right). \end{aligned}$$

For arbitrary tensors A^{ab} and C^{cd} and the corresponding decomposition into self dual / anti-self dual parts we have

$$\varepsilon_{abcd}A^{ab}_{\text{selfdual}}C^{cd}_{\text{antiselfdual}} = \varepsilon_{abcd}A^{ab}_{\text{antiselfdual}}C^{cd}_{\text{selfdual}} = 0,$$

which is easily verified by a short calculation. Therefore we may write the action as

$$\begin{aligned} S_{EH} = -\frac{c^3}{32\pi G}\int \varepsilon_{abcd}\Bigg[&\left(B^{ab}_{\text{selfdual}} \wedge R^{cd}_{\text{selfdual}} + \frac{\Lambda}{6}B^{ab}_{\text{selfdual}} \wedge B^{cd}_{\text{selfdual}}\right) \\ &+ \left(B^{ab}_{\text{antiselfdual}} \wedge R^{cd}_{\text{antiselfdual}} + \frac{\Lambda}{6}B^{ab}_{\text{antiselfdual}} \wedge B^{cd}_{\text{antiselfdual}}\right)\Bigg]. \end{aligned}$$

Within the Plebanski formalism we now postulate that gravitation is determined by the self dual forms alone or equivalently that the anti-self dual forms are vanishing

$$B^{ab}_{\text{antiselfdual}} = 0, \qquad R^{ab}_{\text{antiselfdual}} = 0.$$

With this assumption the action simplifies to

$$S_{EH} = -\frac{c^3}{32\pi G}\int \varepsilon_{abcd}\left(B^{ab}_{\text{selfdual}} \wedge R^{cd}_{\text{selfdual}} + \frac{\Lambda}{6}B^{ab}_{\text{selfdual}} \wedge B^{cd}_{\text{selfdual}}\right).$$

Remark: Within the Plebanski formalism we complexified the tangent space and the cotangent space. The conditions $B^{ab}_{\text{antiselfdual}} = R^{ab}_{\text{antiselfdual}} = 0$ basically define how we continue the differential forms from the real subspace to the complex space.

Remark: Up to now we considered within the Plebanski formalism the vielbein and the spin connection as the fundamental fields. B^{ab}_{selfdual} is constructed out of the vielbein, $R^{ab}_{\text{antiselfdual}}$ is constructed out of the spin connection. It is possible to change the field variables from the vielbein $e_a{}^{\mu}$ to B^{ab}_{selfdual}. However, we have to take care of the correct degrees of freedom. A real vielbein has 16 degrees of freedom, a complex vielbein has 32 degrees of freedom and a complex vielbein with 16 constraints originating from $B^{ab}_{\text{antiselfdual}} = 0$ has again 16 degrees of freedom. On the other hand, if we consider a complex two-form

$$B^{ab} = B^{ab}{}_{\mu\nu}dx^{\mu} \wedge dx^{\nu}$$

with ${B^{ab}}_{\mu\nu}$ anti-symmetric in a,b and μ,ν, we have $2 \cdot 6 \cdot 6 = 72$ degrees of freedom, the self-duality condition reduces this number to 36 degrees of freedom. Thus we have to eliminate $36 - 16 = 20$ degrees of freedoms. The constraints eliminating these degrees of freedom are called **simplicity constraints** and can be implemented by adding a term

$$-\frac{c^3}{32\pi G} \int \psi_{abcd} B^{ab}_{\text{selfdual}} \wedge B^{cd}_{\text{selfdual}}$$

with a Lagrange multiplier field ψ_{abcd} satisfying

$$\psi_{abcd} \;=\; -\psi_{bacd} \;=\; -\psi_{abdc} \;=\; \psi_{cdab}$$

and

$$\varepsilon^{abcd}\psi_{abcd} \;=\; 0.$$

The auxiliary Lagrange multiplier field ψ_{abcd} has the same symmetries as the Riemann curvature tensor and therefore 20 independent components in four space-time dimensions. Variation with respect to ψ_{abcd} gives the twenty simplicity constraints.

7 Special solutions of Einstein's equations

7.1 The Schwarzschild solution

We consider a static spherically symmetric mass distribution, as for example given to a good approximation by the earth or the sun. We are interested in a solution of Einstein's equations outside the mass distribution. Thus we seek solutions of

$$Ric_{\mu\nu} = 0$$

Remark: Einstein's equations (without a cosmological constant) can be written as

$$Ric_{\mu\nu} = \frac{8\pi G}{c^4}\left(T_{\mu\nu} - \frac{1}{2}g_{\mu\nu}g^{\rho\sigma}T_{\rho\sigma}\right).$$

In vacuum we have $T_{\mu\nu} = 0$, hence Einstein's equations reduce to $Ric_{\mu\nu} = 0$.

Remark: The exact definition of "static" and "spherically symmetric" requires some care, as we have to keep coordinate independence. We postpone a detailed discussion. For the moment, let us note that "static" implies that all metric components are time independent and that no mixed terms

$$cdt \otimes dx^i + dx^i \otimes cdt$$

appear in the metric. The last condition can be understood, if we assume that "static" also implies invariance under time reversal $t \to -t$. Under this transformation the terms c^2dt^2 or dx^idx^j don't change their sign, however the mixed terms $cdtdx^i$ do change sign.

Spherical symmetry implies that the infinitesimal solid angle element $d\Omega^2$ does not change its form: The coefficient of the term $d\varphi^2$ should always be $\sin^2\vartheta$ times the coefficient of the term $d\vartheta^2$. Furthermore it implies that there are except for the terms $d\varphi^2$ and $d\vartheta^2$ no further terms (i.e. mixed terms) containing $d\varphi$ or $d\vartheta$.

We make the following ansatz:

$$ds^2 = e^{2a(r)}c^2dt^2 - e^{2b(r)}dr^2 - e^{2c(r)}r^2d\Omega^2$$

We may slightly simplify the ansatz as follows: If we change to a new radial variable defined by

$$r' = e^{c(r)}r$$

we obtain

$$ds^2 = e^{2a(r)}c^2dt^2 - \left(1 + r\frac{dc(r)}{dr}\right)^{-2} e^{2b(r)-2c(r)}dr^2 - r^2d\Omega^2.$$

Thus we see that by a redefinition of the function $b(r)$ it is sufficient to consider the ansatz

$$ds^2 = e^{2a(r)}c^2dt^2 - e^{2b(r)}dr^2 - r^2d\Omega^2.$$

$a(r)$ and $b(r)$ are two functions, which we have to determine. We first compute the Christoffel symbols (and set within the calculation for simplicity $c = 1$):

$$\begin{array}{lll} \Gamma^t{}_{tr} = \partial_r a, & \Gamma^r{}_{tt} = e^{2(a-b)}\partial_r a, & \Gamma^r{}_{rr} = \partial_r b, \\ \Gamma^\theta{}_{r\theta} = \frac{1}{r}, & \Gamma^r{}_{\theta\theta} = -re^{-2b}, & \Gamma^\varphi{}_{r\varphi} = \frac{1}{r}, \\ \Gamma^r{}_{\varphi\varphi} = -re^{-2b}\sin^2\theta, & \Gamma^\theta{}_{\varphi\varphi} = -\sin\theta\cos\theta, & \Gamma^\varphi{}_{\theta\varphi} = \frac{\cos\theta}{\sin\theta}. \end{array}$$

All other components are either related to the ones above by symmetry or are zero. In the next step we compute the components of Riemann's curvature tensor:

$$\begin{array}{lll} R^t{}_{rtr} = a'b' - a'' - (a')^2, & R^t{}_{\theta t\theta} = -re^{-2b}a', & R^t{}_{\varphi t\varphi} = -re^{-2b}\sin^2\theta\, a', \\ R^r{}_{\theta r\theta} = re^{-2b}b', & R^r{}_{\varphi r\varphi} = re^{-2b}\sin^2\theta\, b', & R^\theta{}_{\varphi\theta\varphi} = \left(1 - e^{-2b}\right)\sin^2\theta, \end{array}$$

where we used the notation $a' = \partial_r a$ and $b' = \partial_r b$. We therefore obtain for the components of the Ricci tensor

$$\begin{aligned} Ric_{tt} &= e^{2(a-b)}\left[a'' + (a')^2 - a'b' + \frac{2}{r}a'\right], \\ Ric_{rr} &= -a'' - (a')^2 + a'b' + \frac{2}{r}b', \\ Ric_{\theta\theta} &= e^{-2b}\left[r\left(b' - a'\right) - 1\right] + 1, \\ Ric_{\varphi\varphi} &= \sin^2\theta\, R_{\theta\theta}. \end{aligned}$$

The scalar curvature is given by

$$R = -2e^{-2b}\left[a'' + (a')^2 - a'b' + \frac{2}{r}\left(a' - b'\right) + \frac{1}{r^2}\left(1 - e^{2b}\right)\right].$$

Outside the mass distribution we have

$$Ric_{\mu\nu} = 0.$$

Since Ric_{tt} and Ric_{rr} have to vanish independently we also have

$$0 = e^{2(b-a)}R_{tt} + R_{rr} = \frac{2}{r}\left(a' + b'\right)$$

and therefore $a' + b' = 0$. Integration of this equation leads to

$$b(r) = -a(r) + c.$$

We may eliminate the integration constant c by a rescaling of the time coordinate

$$t \rightarrow e^{-c} t.$$

Hence, we may assume without loss of generality that

$$b(r) = -a(r).$$

We now consider $R_{\theta\theta} = 0$. Substituting the expression for $b(r)$ one obtains

$$e^{2a}\left(2ra' + 1\right) = 1.$$

We may re-write this equation as

$$\frac{d}{dr}\left(re^{2a(r)}\right) = 1.$$

This equation is solved by

$$e^{2a(r)} = 1 - \frac{r_s}{r},$$

as one easily verifies by differentiation. r_s is a yet to be determined integration constant. If we re-insert all factors of the speed of light c, we obtain for the **metric** the result

$$ds^2 = \left(1 - \frac{r_s}{r}\right) c^2 dt^2 - \frac{dr^2}{1 - \frac{r_s}{r}} - r^2 \left(d\theta^2 + \sin^2\theta d\phi^2\right).$$

This solution was found by K. Schwarzschild in 1916. In order to determine the integration constant r_s we study at the tt-component of the metric. For a point mass m we obtain in the Newtonian limit

$$g_{tt} = c^2\left(1 + \frac{2}{c^2}\Phi\right) = c^2\left(1 - \frac{2Gm}{rc^2}\right)$$

and therefore r_s is given by

$$r_s = \frac{2Gm}{c^2}.$$

The quantity r_s is known as the **Schwarzschild radius** of the mass m.

Examples for the Schwarzschild radius:

$$\begin{aligned} \text{Sun}: \; m \approx 2 \cdot 10^{30}\,\text{kg} &\rightarrow r_s = 2.95\,\text{km}, \\ \text{Earth}: \; m \approx 6 \cdot 10^{24}\,\text{kg} &\rightarrow r_s = 0.9\,\text{cm}. \end{aligned}$$

A theorem by Birkhoff states that the Schwarzschild solution is the unique spherically symmetric solution of Einstein's equations in the vacuum. This theorem implies in particular that there are no time-dependent solutions. We sketch a proof of Birkhoff's theorem: We start with the exact definition of "spherical symmetry": In a flat three-dimensional space spherical symmetry corresponds to invariance under the rotation group $SO(3)$. On an arbitrary semi-Riemannian manifold symmetries are characterised by Killing vector fields. The Killing vector fields of the surface of the sphere S^2 are given by

$$\begin{aligned} R &= \partial_\varphi, \\ S &= \cos\varphi\,\partial_\theta - \cot\theta\sin\varphi\,\partial_\varphi, \\ T &= -\sin\varphi\,\partial_\theta - \cot\theta\cos\varphi\,\partial_\varphi. \end{aligned}$$

These vector fields satisfy the commutation relations

$$[R,S] = T, \quad [S,T] = R, \quad [T,R] = S.$$

This is nothing else than the Lie algebra of the group $SO(3)$. We are now in a position to define the concept of "spherical symmetry" for an arbitrary four-dimensional space-time: We require the existence of three Killing vector fields, which satisfy the commutation relations stated above. By a suitable choice of coordinates this implies that the metric can be brought into the form

$$ds^2 = e^{2a(t,r)}c^2dt^2 - e^{2b(t,r)}dr^2 - r^2d\Omega^2.$$

Remark: The functions $a(t,r)$ and $b(t,r)$, which appear in this expression, are a priori functions of t and r. From the form above we may (analogously to what we did before) calculate the Christoffel symbols, the curvature tensor and the Ricci tensor. For example, we find

$$Ric_{tr} = \frac{2}{r}\partial_t b$$

and hence

$$b = b(r).$$

With the help of a suitable coordinate re-definition of the time coordinate we may in addition ensure that $a(t,r)$ does not depend on t. This leads to the ansatz

$$ds^2 = e^{2a(r)}c^2dt^2 - e^{2b(r)}dr^2 - r^2d\Omega^2,$$

which was used for the derivation of the Schwarzschild solution.

Remark: All components of the metric are time-independent. This implies that every spherically symmetric solution of Einstein's equation in the vacuum possesses a time-like Killing vector field.

We call a metric which possesses a Killing vector field that is time-like at infinity a **stationary metric**. The general form of a stationary metric is is given by

$$ds^2 = g_{00}(\vec{x})dt^2 + g_{0i}(\vec{x})\left(dtdx^i + dx^idt\right) + g_{ij}(\vec{x})dx^idx^j.$$

We call a metric which possesses a Killing vector field that is time-like and orthogonal to a family of hypersurfaces a **static metric**. The general form of a static metric is is given by

$$ds^2 = g_{00}(\vec{x})dt^2 + g_{ij}(\vec{x})dx^i dx^j.$$

Let us now consider the singularities of the Schwarzschild metric:

- The metric is singular at $r = r_s$. However this is just a coordinate singularity, physical quantities like the Einstein tensor or the curvature tensor are finite at $r = r_s$. The physical interpretation of $r = r_s$ is given as the event horizon of a black hole.

  Remark: A trivial example for a coordinate singularity is given at the origin of a two-dimensional plane, if one uses polar coordinates:

  $$ds^2 = dr^2 + r^2 d\varphi^2.$$

  $$g_{\mu\nu} = \begin{pmatrix} 1 & 0 \\ 0 & r^2 \end{pmatrix}, \qquad g^{\mu\nu} = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{r^2} \end{pmatrix}.$$

  In particular we have

  $$g^{\varphi\varphi} = \frac{1}{r^2}.$$

  Obviously this is an artefact of the chosen coordinate system, since in a flat plane there are no distinguished points.

- The point $r = 0$ is a proper singularity. In order to distinguish proper singularities from coordinate singularities we consider scalar quantities, like for example

  $$R = g^{\mu\nu} Ric_{\mu\nu}, \quad Ric^{\mu\nu} Ric_{\mu\nu}, \quad R^{\mu\nu\rho\sigma} R_{\mu\nu\rho\sigma}.$$

  For example, one finds for the Schwarzschild metric

  $$R^{\mu\nu\rho\sigma} R_{\mu\nu\rho\sigma} = \frac{12 r_s^2}{r^6}.$$

7.2 The perihelion precession of Mercury

We first consider geodesics for the Schwarzschild metric:

$$\frac{d^2 x^\mu}{d\lambda^2} + \Gamma^\mu_{\tau\sigma} \frac{dx^\tau}{d\lambda} \frac{dx^\sigma}{d\lambda} = 0.$$

The Christoffel symbols for the Schwarzschild metric read (we set again $c = 1$):

$$\begin{array}{lll} \Gamma^t_{tr} = \frac{r_s}{2r(r-r_s)}, & \Gamma^r_{tt} = \frac{r_s}{2r^3}(r - r_s), & \Gamma^r_{rr} = -\frac{r_s}{2r(r-r_s)}, \\ \Gamma^\theta_{r\theta} = \frac{1}{r}, & \Gamma^r_{\theta\theta} = -(r - r_s), & \Gamma^\varphi_{r\varphi} = \frac{1}{r}, \\ \Gamma^r_{\varphi\varphi} = -(r - r_s)\sin^2\theta, & \Gamma^\theta_{\varphi\varphi} = -\sin\theta\cos\theta, & \Gamma^\varphi_{\theta\varphi} = \frac{\cos\theta}{\sin\theta}. \end{array}$$

The geodesic equation gives four coupled second-order differential equations, which are rather difficult to solve directly. A simpler way to the solution proceeds as follows: We already know that the Schwarzschild metric possesses four Killing vector fields: One vector field corresponds to the invariance under time translations, three further vector fields correspond to the spherical symmetry. For a Killing vector field K^μ we have

$$K_\mu \frac{dx^\mu}{d\lambda} = \text{const.}$$

In addition there is one further conserved quantity:

$$\varepsilon = g_{\mu\nu} \frac{dx^\mu}{d\lambda} \frac{dx^\nu}{d\lambda}.$$

For the choice $\lambda = s$ we obtain $\varepsilon = 1$.

The time-like Killing vector field corresponds to energy conservation and is given in the coordinates (t, r, θ, φ) by

$$K^\mu = (\partial_t)^\mu = (1, 0, 0, 0).$$

Lowering the index yields

$$K_\mu = \left(\left(1 - \frac{r_s}{r}\right), 0, 0, 0 \right).$$

The three Killing vector fields associated to the spherical symmetry correspond to the conservation of angular momentum. One vector field corresponds to the magnitude of the angular momentum, two vector fields to the direction of the angular momentum. Conservation of the direction of the angular momentum implies that the particle moves in a plane. We may therefore choose a coordinate system such that the motion of the particle is within the plane defined by

$$\theta = \frac{\pi}{2}.$$

The Killing vector field corresponding to the magnitude of the angular momentum is given by

$$R^\mu = \left(\partial_\varphi\right)^\mu = (0, 0, 0, 1).$$

Lowering the index yields

$$R_\mu = \left(0, 0, 0, -r^2 \sin^2\theta\right).$$

With $\sin\theta = 1$ we have for the conserved quantities

$$\begin{aligned} E &= K_\mu \frac{dx^\mu}{d\lambda} = \left(1 - \frac{r_s}{r}\right) \frac{dt}{d\lambda}, \\ L &= -R_\mu \frac{dx^\mu}{d\lambda} = r^2 \frac{d\varphi}{d\lambda}. \end{aligned}$$

Let us now consider

$$\varepsilon = g_{\mu\nu} \frac{dx^\mu}{d\lambda} \frac{dx^\nu}{d\lambda}.$$

Explicitly, we have

$$\varepsilon = \left(1 - \frac{r_s}{r}\right) \left(\frac{dt}{d\lambda}\right)^2 - \left(1 - \frac{r_s}{r}\right)^{-1} \left(\frac{dr}{d\lambda}\right)^2 - r^2 \left(\frac{d\varphi}{d\lambda}\right)^2.$$

We substitute the expressions for the conserved quantities E and L and obtain

$$\left(\frac{dr}{d\lambda}\right)^2 + \left(1 - \frac{r_s}{r}\right) \left(\varepsilon + \frac{L^2}{r^2}\right) = E^2.$$

This equation may be written as

$$\frac{1}{2} \left(\frac{dr}{d\lambda}\right)^2 + V(r) = \mathscr{E},$$

with

$$\begin{aligned} \mathscr{E} &= \frac{1}{2} E^2, \\ V(r) &= \frac{1}{2}\varepsilon - \frac{\varepsilon r_s}{2r} + \frac{L^2}{2r^2} - \frac{r_s L^2}{2r^3}. \end{aligned}$$

Remark: Within the Newtonian theory we would find an effective potential which does not include the $1/r^3$-term, but is otherwise identical. The first term of the effective potential is a constant, the second term corresponds to the Newtonian gravitational potential, the third term gives a contribution due to the angular momentum. The form of this term is identical within Newtonian mechanics and general relativity. The last term appears only within general relativity.

The planets move along ellipses around the sun. The point of closest distance to the sun is called the perihelion. Let us now consider the perihelion precession of Mercury. To this aim we determine an equation, which gives the radial coordinate r as a function of the angle φ, i.e. $r = r(\varphi)$. We multiply the equation of motion with

$$\left(\frac{d\varphi}{d\lambda}\right)^{-2} = \frac{r^4}{L^2}$$

and obtain

$$\left(\frac{dr}{d\varphi}\right)^2 + \frac{\varepsilon}{L^2} r^4 - \frac{\varepsilon r_s}{L^2} r^3 + r^2 - r_s r = \frac{E^2}{L^2} r^4.$$

We set

$$x = \frac{2L^2}{r_s r}$$

and obtain

$$\left(\frac{dx}{d\varphi}\right)^2 + 4\frac{L^2}{r_s^2}\left(\varepsilon - E^2\right) - 2\varepsilon x + x^2 = \frac{1}{2}\frac{r_s^2}{L^2}x^3.$$

Let us now differentiate with respect to φ:

$$2\frac{dx}{d\varphi}\frac{d^2x}{d\varphi^2} - 2\varepsilon\frac{dx}{d\varphi} + 2x\frac{dx}{d\varphi} = \frac{3}{2}\frac{r_s^2}{L^2}x^2\frac{dx}{d\varphi}.$$

We obtain the following equation:

$$\frac{d^2x}{d\varphi^2} - \varepsilon + x = \frac{3}{4}\frac{r_s^2}{L^2}x^2.$$

We recall that the parameter ε was defined by

$$\varepsilon = g_{\mu\nu}\frac{dx^\mu}{d\lambda}\frac{dx^\nu}{d\lambda}.$$

If we choose as curve parameter λ the proper time s, i.e. $\lambda = s$, we have

$$g_{\mu\nu}\frac{dx^\mu}{ds}\frac{dx^\nu}{ds} = 1.$$

With $\varepsilon = 1$ our equation reads

$$\frac{d^2x}{d\varphi^2} - 1 + x = \frac{3}{4}\frac{r_s^2}{L^2}x^2.$$

Within Newtonian mechanics the term on the right-hand side is absent and the equation

$$\frac{d^2x}{d\varphi^2} - 1 + x = 0.$$

may be solved exactly:

$$x_{Newton}(\varphi) = 1 + e\cos\varphi.$$

This is the solution of Kepler and Newton and describes a perfect ellipse. The quantity e gives the eccentricity of the ellipse. Within general relativity we treat the term

$$\frac{3}{4}\frac{r_s^2}{L^2}x^2$$

as a small perturbation and seek a solution of the form

$$x(\varphi) = x_{Newton}(\varphi) + \tilde{x}(\varphi).$$

Within perturbation theory we obtain for $\tilde{x}$ the differential equation

$$\begin{aligned}\frac{d^2\tilde{x}}{d\varphi^2}+\tilde{x} &= \frac{3}{4}\frac{r_s^2}{L^2}x_{Newton}^2 \\ &= \frac{3}{4}\frac{r_s^2}{L^2}\left(1+e\cos\varphi\right)^2 \\ &= \frac{3}{4}\frac{r_s^2}{L^2}\left[\left(1+\frac{e^2}{2}\right)+2e\cos\varphi+\frac{e^2}{2}\cos 2\varphi\right].\end{aligned}$$

We have

$$\begin{aligned}\frac{d^2}{d\varphi^2}\left(\varphi\sin\varphi\right)+\varphi\sin\varphi &= 2\cos\varphi, \\ \frac{d^2}{d\varphi^2}\left(\cos 2\varphi\right)+\cos 2\varphi &= -3\cos 2\varphi.\end{aligned}$$

It follows that

$$\tilde{x} = \frac{3}{4}\frac{r_s^2}{L^2}\left[\left(1+\frac{e^2}{2}\right)+e\varphi\sin\varphi-\frac{e^2}{6}\cos 2\varphi\right]$$

is a solution. The first term $1+e^2/2$ corresponds to a constant displacement of x (respectively r), the third term $-e^2/6\cos 2\varphi$ represents an oscillation, which averages to zero. Of particular interest is the second term $e\varphi\sin\varphi$, which accumulates over successive orbits. We neglect the first and the third term and obtain

$$x(\varphi) = 1+e\cos\varphi+\frac{3}{4}\frac{r_s^2}{L^2}e\varphi\sin\varphi.$$

Approximatively we have

$$\cos\left((1-\alpha)\varphi\right) \approx \cos\varphi+\alpha\frac{d}{d\alpha}\cos\left((1-\alpha)\varphi\right)|_{\alpha=0} = \cos\varphi+\alpha\varphi\sin\varphi$$

and therefore

$$x(\varphi) = 1+e\cos\left((1-\alpha)\varphi\right),$$

with

$$\alpha = \frac{3}{4}\frac{r_s^2}{L^2}.$$

We see that the perihelion advances per orbit by an angle

$$\Delta\varphi = 2\pi\alpha = \frac{3\pi r_s^2}{2L^2}.$$

Let us determine L^2: For a perfect ellipse we have

$$r = \frac{(1-e^2)a}{1+e\cos\varphi},$$

where a denotes the semi-major axis. On the other hand we have with $x = 1+e\cos\varphi$

$$r = \frac{2L^2}{r_s}\frac{1}{1+e\cos\varphi}$$

and hence

$$L^2 = \frac{r_s}{2}(1-e^2)a.$$

With $r_s = 2Gm/c^2$ we finally obtain

$$\Delta\varphi = \frac{6\pi Gm}{(1-e^2)a}.$$

For the sun we have

$$\frac{Gm}{c^2} = 1.48\cdot 10^3\text{m}.$$

The orbit of Mercury is specified by

$$a = 5.79\cdot 10^{10}\text{m}, \qquad e = 0.2056.$$

We therefore find

$$\Delta\varphi = 0.103''/\text{orbit}.$$

The precession is usually quoted per century. The time for one orbit for Mercury is 88 days. We therefore find

$$\Delta\varphi = 43.0''/(100\text{ y}).$$

We may now compare this number to the observed value:

$$\begin{array}{rl} 5601''/(100\text{ y}) & \text{measured} \\ -5025''/(100\text{ y}) & \text{precession of equinoxes} \\ \underline{-532''/(100\text{ y})} & \text{perturbation due to other planets} \\ 44''/(100\text{ y}) & \end{array}$$

The primary data are optical positions of Mercury on the sky as measured from the earth. We have to take into account an apparent perihelion shift caused by the precession of the Earth's rotational axis. This is called the precession of the equinoxes and is related to the angle of 23.5° of the Earth's equatorial plane against the Earth's ecliptic plane (defined by the Earth's motion around the sun).

7.3 Black holes, Kruskal coordinates and Penrose diagrams

In our previous discussion of the Schwarzschild solution we focussed on the exterior region ($r > r_s$). Let us now see what happens as we approach the Schwarzschild radius r_s. We start by studying the causal structure. We consider light rays for constant θ and φ:

$$ds^2 = 0 = \left(1 - \frac{r_s}{r}\right) c^2 dt^2 - \frac{dr^2}{1 - \frac{r_s}{r}}.$$

Therefore

$$\frac{cdt}{dr} = \pm \frac{1}{1 - \frac{r_s}{r}}.$$

For large r the right-hand side approaches ± 1, however for $r \to r_s$ we find

$$\lim_{r \to r_s} \frac{cdt}{dr} = \pm \infty.$$

In this coordinate system the light cones become narrower as we approach the Schwarzschild radius. This does not mean that it is impossible to cross the Schwarzschild radius. An object has no problems moving towards the black hole. If the object emits in regular intervals (with respect to the object's proper time) light signals, an observer on Earth will receive these light signals with increasing gaps in-between. The observer on Earth will only receive the signals, which were emitted before the crossing of the Schwarzschild radius.

In order to understand better the event horizon at r_s we try to find better coordinate systems which do not possess a coordinate singularity at $r = r_s$. We will do this in several steps. Let us define for $r > r_s$

$$r^* = r + r_s \ln\left(\frac{r}{r_s} - 1\right).$$

The metric reads now

$$ds^2 = \left(1 - \frac{r_s}{r}\right)\left(c^2 dt^2 - dr^{*2}\right) - r^2 d\Omega^2,$$

where r should now be understood as a function of r^*. We now have

$$\frac{cdt}{dr^*} = \pm 1,$$

however the event horizon $r = r_s$ corresponds now to $r^* = -\infty$. If we define

$$\begin{aligned} v &= ct + r^*, \\ u &= ct - r^*, \end{aligned}$$

we see that infalling radial light-like geodesics are characterised by $v = const$, while outgoing radial light-like geodesics are characterised by $u = const$.

If we now go back to the original radial coordinate r, but replace the time coordinate by

$$v = ct + r^* = ct + r + r_s \ln\left(\frac{r}{r_s} - 1\right),$$

we obtain coordinates known as **Eddington-Finkelstein coordinates**. The metric reads in these coordinates

$$ds^2 = \left(1 - \frac{r_s}{r}\right) dv^2 - (dvdr + drdv) - r^2 d\Omega^2.$$

The determinant of the metric is given in these coordinates by

$$\begin{vmatrix} 1 - \frac{r_s}{r} & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & -r^2 & 0 \\ 0 & 0 & 0 & -r^2 \sin^2\theta \end{vmatrix} = -r^4 \sin^2\theta.$$

The determinant does not have a singularity for $r = r_s$. For radial light-like curves we have

$$\frac{dv}{dr} = \begin{cases} 0, & \text{always infalling} \\ \frac{2}{1-\frac{r_s}{r}}, & \text{outgoing for } r > r_s, \text{infalling for } r < r_s, \end{cases}$$

At $r = r_s$ a radially outgoing ray turns into an infalling ray. We see that $r = r_s$ is a point of no return: If a particle crosses $r = r_s$ it will never return. We define the **event horizon** as the surface beyond which particles can never return to spatial infinity. The region bound by the event horizon is called a **black hole**.

Up to now we found for the Schwarzschild space-time two regions: the exterior region $r > r_s$ and the region of the black hole, which can be reached from the exterior region on future-directed curves. Let us note that it is impossible to reach the black hole on past-directed curves.

The Schwarzschild solution is static and therefore invariant under time reversal. Therefore the two regions found up to now cannot constitute the complete space-time. A further region is obtained if we use in the redefinition of the time coordinate instead of v the variable u:

$$u = ct - r^* = ct - r - r_s \ln\left(\frac{r}{r_s} - 1\right),$$

The metric reads now

$$ds^2 = \left(1 - \frac{r_s}{r}\right) du^2 - (dudr + drdu) - r^2 d\Omega^2.$$

In this coordinate system the region $r < r_s$ is a region which can be reached from the exterior region on past-directed curves, but never on future-directed curves. Signals from this region may reach the exterior region. However, it is impossible for particles to reach this region. This region is called a **white hole**.

In order to cover all regions of the Schwarzschild space-time with a single coordinate system, we introduce a new coordinate system through

$$\begin{aligned} T &= \sqrt{\frac{r}{r_s}-1}\, e^{\frac{r}{2r_s}}\ \sinh\left(\frac{ct}{2r_s}\right), \\ R &= \sqrt{\frac{r}{r_s}-1}\, e^{\frac{r}{2r_s}}\ \cosh\left(\frac{ct}{2r_s}\right). \end{aligned}$$

The metric reads in these coordinates

$$ds^2 = \frac{4r_s^3}{r} e^{\frac{r}{r_s}} \left(dT^2 - dR^2\right) - r^2 d\Omega^2,$$

where now r is implicitly defined through

$$T^2 - R^2 = \left(1 - \frac{r}{r_s}\right) e^{\frac{r}{r_s}}.$$

The coordinates (T, R, θ, φ) are known as **Kruskal coordinates**. In Kruskal coordinates we have for radial light-like curves

$$T = \pm R + const.$$

The event horizon $r = r_s$ is given by

$$T = \pm R.$$

More generally, we have for surfaces defined by $r = const$:

$$T^2 - R^2 = const.$$

The allowed regions of (T, R) are therefore given by

$$-\infty \leq R \leq \infty, \qquad T^2 < R^2 + 1.$$

Surfaces defined by $t = const$ are given by

$$\frac{T}{R} = \tanh\left(\frac{ct}{2r_s}\right).$$

The essential properties of a space-time can represented with the help of a **Penrose diagram**. Penrose diagrams have the following properties:

- Penrose diagrams display the time coordinate and the radial coordinate.
- Light rays in radial direction are in Penrose diagrams lines at angles of 45°.
- Penrose diagrams represent the entire space-time in a finite region.

Let us first consider the construction of the Penrose diagram for the flat Minkowski space-time. We start with the metric in spherical coordinates (and for simplicity we set $c = 1$):

$$ds^2 = dt^2 - dr^2 - r^2 d\Omega^2.$$

We then define light-cone coordinates

$$u = t - r, \qquad v = t + r.$$

The regions of u and v are:

$$-\infty < u < \infty, \quad -\infty < v < \infty, \quad u \le v.$$

The metric reads now

$$ds^2 = \frac{1}{2}\left(dudv + dvdu\right) - \frac{1}{4}\left(v-u\right)^2 d\Omega^2.$$

Let us set

$$\begin{aligned} u' &= \arctan u, \\ v' &= \arctan v. \end{aligned}$$

The allowed region transforms to

$$-\frac{\pi}{2} < u' < \frac{\pi}{2}, \qquad -\frac{\pi}{2} < v' < \frac{\pi}{2}, \qquad u' \le v'.$$

The metric is now given by

$$ds^2 = \frac{1}{4\cos^2 u' \cos^2 v'}\left[2\left(du'dv' + dv'du'\right) - \sin^2(v'-u')d\Omega^2\right].$$

Finally we set

$$\begin{aligned} t' &= v' + u', \\ r' &= v' - u'. \end{aligned}$$

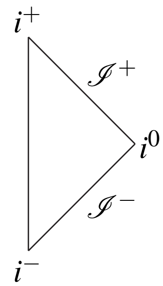
This gives the region

$$0 \le r' < \pi, \qquad \left|t'\right| + r' < \pi,$$

and the metric

$$ds^2 = \frac{1}{\left(\cos t' + \cos r'\right)^2}\left(dt'^2 - dr'^2 - \sin^2 r' d\Omega^2\right].$$

The Penrose diagram of Minkowski space-time:



Within a Penrose diagram one denotes by

i^+	infinity for all future-directed time-like curves,
i^0	infinity for all space-like curves,
i^-	infinity for all past-directed time-like curves,
$\mathscr{I}^+$	infinity for future-directed light-like curves,
$\mathscr{I}^-$	infinity for all past-directed light-like curves.

All time-like geodesics start at i^- and end at i^+. All space-like geodesics start and end at i^0. All light-like geodesics start at $\mathscr{I}^-$ and end at $\mathscr{I}^+$. (Light rays, which start at $\mathscr{I}^-$ are first radially incoming until $r = 0$. Afterwards they are radially outgoing. If we draw such a light ray in a Penrose diagram, it is effectively reflected at $r = 0$. This light ray ends at $\mathscr{I}^+$.)

We call a space-time (or a region of a space-time) **asymptotically flat**, if in the associated Penrose diagram $\mathscr{I}^+$, i^0 and $\mathscr{I}^-$ are as in the Penrose diagram of Minkowski space-time.

The Penrose diagram of the Scharzschild space-time is obtained in along the same lines. We start with the Kruskal coordinates and define

$$U = T - R, \qquad V = T + R.$$

We then set

$$U' = \arctan \frac{U}{\sqrt{r_s}}, \qquad V' = \arctan \frac{V}{\sqrt{r_s}},$$

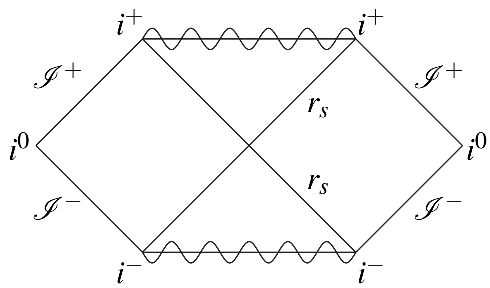
and finally

$$T' = V' + U', \qquad R' = V' - U'.$$

The region is given by

$$-\frac{\pi}{2} < U' < \frac{\pi}{2}, \quad -\frac{\pi}{2} < V' < \frac{\pi}{2}, \quad -\frac{\pi}{2} < U' + V' < \frac{\pi}{2}.$$

The Penrose diagram of the Schwarzschild space-time:



7.4 Charged black holes: The Reissner-Nordström solution

There is a theorem which states that within general relativity coupled to electrodynamics stationary, asymptotically flat black hole solutions, which are non-singular outside an event horizon are completely characterised by the three quantities mass, charge and angular momentum. This theorem is known as the **no-hair theorem**.

The Schwarzschild solution corresponds to the case, where the charge and the angular momentum are zero. Let us now generalise this solution to the case, where we allow a non-zero charge (but still take the angular momentum to be zero). The Reissner-Nordström solution describes an electrically charged black hole. The charge of the black hole is denoted by Q. The metric is given by

$$ds^2 = \frac{\Delta}{r^2} c^2 dt^2 - \frac{r^2}{\Delta} dr^2 - r^2 d\Omega^2,$$

where

$$\Delta = r^2 - \frac{2Gmr}{c^2} + \frac{GQ^2}{c^4}.$$

We set $c = 1$. We then obtain

$$\Delta = r^2 - 2Gmr + GQ^2.$$

This solution was worked out in the years 1916-1918 by Reissner and Nordström. The event horizon is obtained from the equation

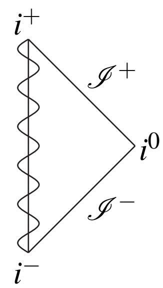
$$\begin{aligned} \Delta &= 0, \\ r_\pm &= Gm \pm \sqrt{G^2 m^2 - GQ^2}. \end{aligned}$$

We consider the following cases:

Case 1: $Gm^2 < Q^2$.
In this case there is no real solution for $r_\pm$.. The quantity Δ is always positive and the metric is

regular for all points $r \neq 0$. There is no event horizon separating the singularity at $r = 0$ from the asymptotically flat region. A singularity from which signals can reach $\mathscr{I}^+$ is called a **naked singularity**.

The Penrose diagram of the Reissner-Nordström solution for $Gm^2 < Q^2$:

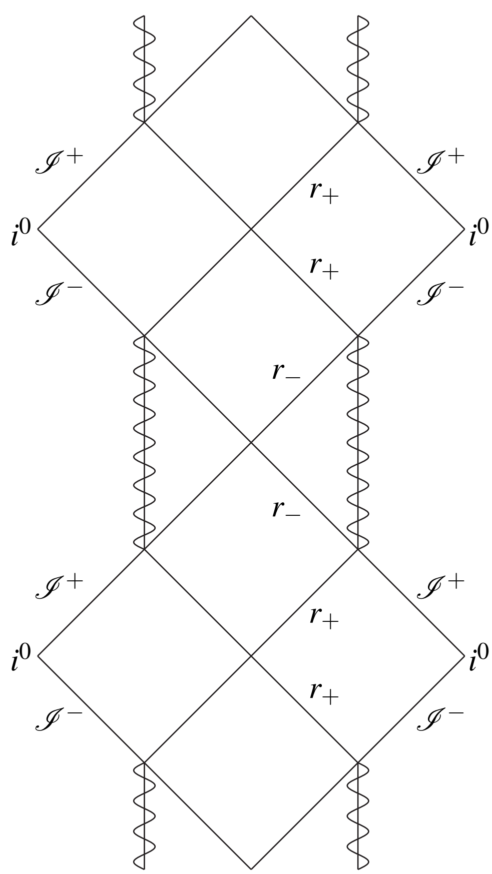


Case 2: $Gm^2 > Q^2$. In this case we have event horizons at

$$r_\pm = Gm \pm \sqrt{G^2m^2 - GQ^2}.$$

The singularity at $r = 0$ is time-like.

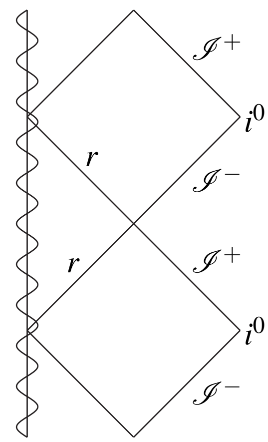
The Penrose diagram of the Reissner-Nordström solution for $Gm^2 > Q^2$:



Case 3: $Gm^2 = Q^2$. This case is called the extreme Reissner-Nordström solution. In this case the values r_+ and r_- coincide:

$$r = Gm.$$

The singularity at $r = 0$ is time-like, if one crosses the event horizon it is possible to avoid the singularity and to enter another asymptotically flat region.
The Penrose diagram of the Reissner-Nordström solution for $Gm^2 = Q^2$:



7.5 Rotating black holes: The Kerr solution

The Kerr solution describes a rotating black hole (with zero electric charge). The angular momentum of the black hole is denoted by J. The metric reads

$$\begin{aligned} ds^2 &= \left(1 - \frac{2Gmr}{c^2\Sigma}\right) c^2 dt^2 + \frac{2Gmr\, j \sin^2\theta}{c^2\Sigma} \left(cdtd\varphi + d\varphi cdt\right) \\ & -\frac{\Sigma}{\Delta} dr^2 - \Sigma d\theta^2 - \left(\frac{(r^2+j^2)^2 - \Delta j^2 \sin^2\theta}{\Sigma}\right) \sin^2\theta d\varphi^2, \end{aligned}$$

where

$$\Delta = r^2 - \frac{2Gmr}{c^2} + j^2, \quad \Sigma = r^2 + j^2\cos^2\theta, \quad j = \frac{J}{mc}.$$

This solution was found by Kerr in 1953.

Let us also consider the most general case: A rotating and electrically charged black hole of mass m, charge Q and angular momentum J. The metric reads

$$\begin{aligned} ds^2 &= \left(\frac{\Delta - j^2\sin^2\theta}{\Sigma}\right) c^2 dt^2 + \frac{j \sin^2\theta \left(r^2 + j^2 - \Delta\right)}{\Sigma} \left(cdtd\varphi + d\varphi cdt\right) \\ & -\frac{\Sigma}{\Delta} dr^2 - \Sigma d\theta^2 - \left(\frac{(r^2+j^2)^2 - j^2 \Delta \sin^2\theta}{\Sigma}\right) \sin^2\theta d\varphi^2, \end{aligned}$$

where

$$\Delta = r^2 - \frac{2Gmr}{c^2} + \frac{GQ^2}{c^4} + j^2, \quad \Sigma = r^2 + j^2 \cos^2\theta, \quad j = \frac{J}{mc}.$$

This metric is known as the **Kerr-Newman metric**.

The coordinates (t, r, θ, φ) are also known as Boyer-Lindquist coordinates. For $Q = J = 0$ the Kerr-Newman metric reduces to the Schwarzschild metric.

We discuss a few peculiarities related to the non-zero angular momentum. For simplicity we consider the original Kerr solution ($Q = 0$). If we keep j constant and then consider the limit $m \to 0$ we obtain

$$ds^2 = c^2dt^2 - \frac{(r^2 + j^2\cos^2\theta)}{r^2 + j^2}dr^2 - (r^2 + j^2\cos^2\theta)d\theta^2 - (r^2 + j^2)\sin^2\theta d\varphi^2.$$

This is the Minkowski metric in ellipsoidal coordinates

$$\begin{aligned} x &= \sqrt{r^2 + j^2}\sin\theta\cos\varphi, \\ y &= \sqrt{r^2 + j^2}\sin\theta\sin\varphi, \\ z &= r\cos\theta. \end{aligned}$$

In particular, $r = 0$ corresponds to a two-dimensional disc.

The Kerr metric is not static, but stationary. The metric contains mixed terms $(cdtd\varphi + d\varphi cdt)$. The event horizon is again given by the solution of the equation (we set again $c = 1$)

$$\Delta = r^2 - 2Gmr + j^2 = 0.$$

As in the case of the Reissner-Nordström solution we distinguish also for the Kerr solution three cases: $Gm < j$, $Gm = j$ and $Gm > j$. We limit ourselves to discuss the last case in more detail. In the case $Gm > j$ we find

$$r_\pm = Gm \pm \sqrt{G^2m^2 - j^2}.$$

Previously we defined the event horizon as a hypersurface beyond which particles can never return to spatial infinity. The event horizon is a light-like hypersurface. We say that a light-like hypersurface Σ is a **Killing horizon** of a Killing vector field K, if K is light-like on Σ. For the Schwarzschild metric and the Reissner-Nordström metric we may consider the Killing vector field $K = \partial_t$. In this case the Killing horizon coincides with the event horizon.
However, this is no longer true for the Kerr metric: The Killing horizon of the vector field $K = \partial_t$ is not identical to the event horizon. The reason is, that the Kerr solution is stationary, but not static. We obtain the Killing horizon of the vector field $K = \partial_t$ by solving the equation $K^\mu K_\mu = 0$. This leads to

$$(r - Gm)^2 = G^2m^2 - j^2\cos^2\theta.$$

Let us compare this equation to the equation satisfied by the outer event horizon r_+:

$$(r_+ - Gm)^2 = G^2m^2 - j^2.$$

The region between these two hypersurfaces is known as **ergosphere**.

8 A brief review of statistical physics

Before we start to discuss cosmology, it is worth to review a few key ingredients of thermodynamics and statistical physics.

The entropy of a system consisting of a single particle species is given by

$$S = \frac{E}{T} + \frac{p \cdot V}{T} - \frac{\mu N}{T},$$

where E denotes the internal energy (usually denoted by U within statistical physics), T the temperature, p the pressure, V the volume, μ the chemical potential and N the particle number.

For a system of bosons, the average occupation number of a state with energy E_i is given by the **Bose-Einstein distribution**

$$\bar{n}_i = \frac{1}{e^{\frac{(E_i-\mu)}{k_B T}} - 1},$$

while for fermions the average occupation number is given by the **Fermi-Dirac distribution**

$$\bar{n}_i = \frac{1}{e^{\frac{(E_i-\mu)}{k_B T}} + 1}.$$

The thermal wavelength λ and the average particle distance l are given by

$$\lambda = \frac{h}{\sqrt{2\pi m k_B T}}, \qquad l = \left(\frac{V}{N}\right)^{\frac{1}{3}}.$$

In the limit where the thermal wavelength is much smaller than the average particle distance ($\lambda \ll l$) both the Bose-Einstein distribution and the Fermi-Dirac distribution reduce to the **Maxwell-Boltzmann distribution**

$$\bar{n}_i = e^{-\frac{(E_i-\mu)}{k_B T}}.$$

It can be shown that the limit $\lambda \ll l$ is equivalent to $z \ll 1$, where

$$z = e^{\frac{\mu}{k_B T}}$$

denotes the **fugacity**.

The number of occupied states in d^3p is

$$g \bar{n}_i V \frac{d^3 p}{(2\pi\hbar)^3},$$

where g denotes the degeneracy factor (the number of spin states). Let us now consider massless particles. With $d^3p = 4\pi p^2 dp$ and $p = \hbar\omega/c$ we obtain for the number of occupied states in $d\omega$

$$g \bar{n}_i \frac{V}{2\pi^2 c^3} \omega^2 d\omega.$$

The **spectral energy density** $u(\omega)$ is defined as energy per volume and unit frequency. We obtain the spectral energy density by multiplying the expression above by $\hbar\omega/V/d\omega$:

$$u(\omega, T) = \frac{g\hbar\omega^3}{2\pi^2 c^3}\bar{n}(\omega).$$

For photons ($g = 2$, $\mu = 0$) we recover Planck's radiation law:

$$u(\omega, T) = \frac{\hbar\omega^3}{\pi^2 c^3}\frac{1}{e^{\frac{\hbar\omega}{k_B T}} - 1}.$$

9 Friedmann-Robertson-Walker cosmology

9.1 Summary on Einstein's equations

A typical problem in electrodynamics is the following: Given a current density j_μ, solve the differential equation

$$\partial^\mu F_{\mu\nu} = \frac{4\pi}{c} j_\nu$$

for A_μ (or $\vec{E}$ and $\vec{B}$). The analogue problem in general relativity is the following: Given an energy-momentum tensor $T_{\mu\nu}$, solve the differential equation

$$Ric_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R - \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu},$$

for $g_{\mu\nu}$. Within cosmology we will assume that $T_{\mu\nu}$ is given.

Let us briefly recall, how the left-hand side of Einstein's equations depends on the metric $g_{\mu\nu}$. We consider a semi-Riemannian manifold (space-time) with the Levi-Civita connection. The connection coefficients are given by the Christoffel symbols

$$\Gamma^\kappa_{\ \mu\nu} = \frac{1}{2} g^{\kappa\lambda} \left(\partial_\mu g_{\nu\lambda} + \partial_\nu g_{\mu\lambda} - \partial_\lambda g_{\mu\nu} \right).$$

From the Christoffel symbols we obtain Riemann's curvature tensor as

$$R^\kappa_{\ \lambda\mu\nu} = \partial_\mu \Gamma^\kappa_{\ \nu\lambda} - \partial_\nu \Gamma^\kappa_{\ \mu\lambda} + \Gamma^\eta_{\ \nu\lambda} \Gamma^\kappa_{\ \mu\eta} - \Gamma^\eta_{\ \mu\lambda} \Gamma^\kappa_{\ \nu\eta}.$$

The Ricci tensor $Ric_{\mu\nu}$ and the scalar curvature R are defined by

$$\begin{aligned} Ric_{\mu\nu} &= R^\lambda_{\ \mu\lambda\nu}, \\ R &= g^{\mu\nu} Ric_{\mu\nu}. \end{aligned}$$

Thus we see that the left-hand side of Einstein's equations depends on the metric and the first and second derivatives thereof.

Let $x^\mu(\lambda)$ be a curve describing the world-line of a free particle. Free particles move in curved space along geodesics, thus

$$\frac{d^2 x^\mu}{d\lambda^2} + \Gamma^\mu_{\ \tau\sigma} \frac{dx^\tau}{d\lambda} \frac{dx^\sigma}{d\lambda} = 0$$

As curve parameter it is convenient to choose for massive particles $\lambda = s/(mc) = \tau/m$, where τ is the proper time of the particle and $s = c\tau$. We then have

$$p^\mu = \frac{dx^\mu}{d\lambda}$$

and the geodesic equation reads

$$\frac{d}{d\lambda}p^{\mu}+\Gamma^{\mu}_{\tau\sigma}p^{\tau}p^{\sigma} = 0.$$

For massless particles we may still normalise the curve parameter such that $p^{\mu}=dx^{\mu}/d\lambda$, yielding the same geodesic equation in terms of momenta.

We further have

$$g_{\mu\nu}p^{\mu}p^{\nu} = m^2c^2.$$

9.2 The perfect fluid

A fluid often gives a good approximation for a system with many particles. Instead of specifying the individual coordinates and velocities of each particle, it is often sufficient to specify just the four-velocity field $u^{\mu}(x)$ of the fluid.

A special role is played by the concept of a **perfect fluid**: By definition, a perfect fluid is described in the rest frame of the fluid by two parameters: the energy density ρ and the pressure density p.

We are in particular interested in the energy-momentum tensor of the perfect fluid. We may motivate the expression for energy-momentum tensor of the perfect fluid as follows: We start in flat space-time and in the rest frame of the fluid. By definition, $T_{\mu\nu}$ depends only on ρ and p:

$$T_{\mu\nu} = \begin{pmatrix} \rho & 0 & 0 & 0 \\ 0 & p & 0 & 0 \\ 0 & 0 & p & 0 \\ 0 & 0 & 0 & p \end{pmatrix}.$$

We now seek a generalisation to coordinate systems related to the rest frame by a Lorentz transformation (recall that we are still in flat space-time). Taking into account that in the rest frame we have

$$u_{\mu}u_{\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad g_{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix},$$

we easily derive the sought-after generalisation:

$$T_{\mu\nu} = (p+\rho)\,u_{\mu}u_{\nu}-pg_{\mu\nu}.$$

In the final step we will assume that this expression is also valid in curved space-time.

The concept of a perfect fluid is rather general and allows to describe a variety of physical situations. In order to specialise to a specific physical situation we impose in addition an **equation of state**

$$p = p(\rho),$$

which gives a relation between the pressure density and the energy density. Examples are

- **Dust**: For dust we have the equation of state

  $$p = 0.$$

  In this case the energy-momentum tensor is given by

  $$T_{\mu\nu} = \rho u_\mu u_\nu.$$

  For example, non-interacting galaxies can be modelled by dust.

- **Photon gas**: For an isotropic photon gas we have the equation of state

  $$p = \frac{1}{3}\rho$$

  and the energy-momentum tensor reduces to

  $$T_{\mu\nu} = \frac{4}{3}\rho u_\mu u_\nu - \frac{1}{3}\rho g_{\mu\nu}.$$

- **Vacuum energy**: Here we have the equation of state

  $$p = -\rho$$

  and the energy-momentum tensor reduces to

  $$T_{\mu\nu} = \rho g_{\mu\nu}.$$

  Remark: If we start from Einstein's equations without a cosmological constant

  $$Ric_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = \frac{8\pi G}{c^4} T_{\mu\nu}$$

  and by decomposing the energy-momentum tensor into a part corresponding to the vacuum energy and a remaining part corresponding to all other matter

  $$T_{\mu\nu} = T_{\mu\nu}^{(M)} + \rho_{vac} g_{\mu\nu}$$

  we obtain

  $$Ric_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R - \frac{8\pi G}{c^4} \rho_{vac} g_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu}^{(M)}.$$

This is equivalent to a cosmological constant

$$\Lambda = \frac{8\pi G}{c^4}\rho_{vac}.$$

Modern cosmology views a term proportional to $g_{\mu\nu}$ in Einstein's equations as a vacuum energy and part of the energy-momentum tensor. Thus Λ is set to zero on the left-hand side and Einstein's equations read

$$Ric_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = \frac{8\pi G}{c^4}T_{\mu\nu},$$

where $T_{\mu\nu}$ includes a vacuum energy. We will adopt this convention from now on.

9.3 Energy conditions

Instead of using specific models for the energy-momentum tensor, it is sometimes useful to discuss in full generality characteristics of solutions of Einstein's equations, which derive from certain properties of the energy-momentum tensor. These properties of the energy-momentum tensor are formulated as energy conditions:

- The **weak energy condition**:

$$T_{\mu\nu}t^\mu t^\nu \geq 0 \quad \text{for all time-like vectors } t^\mu.$$

Applied to the perfect fluid this translates to $\rho \geq 0$ and $\rho + p \geq 0$. These two conditions are obtained as follows: Let us first consider the limit, where the time like vector t^μ approaches a light-like vector l^μ. In this limit we have

$$T_{\mu\nu}l^\mu l^\nu = (p+\rho)(l \cdot u)^2,$$

and hence $p + \rho \geq 0$. Let us then consider the case where $t^\mu = u^\mu$. In this case we have

$$T_{\mu\nu}u^\mu u^\nu = \rho,$$

and hence $\rho \geq 0$. It remains to show that for an arbitrary time-like vector t^μ no other constraints arise. We have

$$\begin{aligned} T_{\mu\nu}t^\mu t^\nu &= (p+\rho)(t\cdot u)^2 - pt^2 = p\left[(t\cdot u)^2 - t^2\right] + \rho\,(t\cdot u)^2 \\ &= p\left[(t\cdot u)^2 - t^2u^2\right] + \rho\,(t\cdot u)^2. \end{aligned}$$

The Schwarz inequality in Lorentzian signature for two time-like four-vectors reads

$$(t\cdot u)^2 - t^2u^2 \geq 0.$$

Using $p \geq -\rho$ we obtain

$$T_{\mu\nu}t^\mu t^\nu \geq -\rho\left[(t\cdot u)^2 - t^2u^2\right] + \rho\,(t\cdot u)^2 = \rho t^2u^2.$$

Using $\rho \geq 0$ it follows

$$T_{\mu\nu} t^\mu t^\nu \geq 0,$$

i.e. no other constraints arise.

- The **null energy condition**:

$$T_{\mu\nu} l^\mu l^\nu \geq 0 \quad \text{for all light-like vectors } l^\mu.$$

  Applied to the perfect fluid this translates to $\rho + p \geq 0$.

- The **dominant energy condition**:

$$\begin{aligned} T_{\mu\nu} t^\mu t^\nu &\geq 0 \quad \text{for all time-like vectors } t^\mu; \\ g^{\mu\nu}\, T_{\mu\rho} t^\rho\, T_{\nu\sigma} t^\sigma &\geq 0, \quad \text{i.e. } T_{\mu\rho} t^\rho \text{ is not space-like.} \end{aligned}$$

  Applied to the perfect fluid this translates to $\rho \geq |p|$

- The **null dominant energy condition**:

$$\begin{aligned} T_{\mu\nu} l^\mu l^\nu &\geq 0 \quad \text{for all light-like vectors } l^\mu; \\ g^{\mu\nu}\, T_{\mu\rho} l^\rho\, T_{\nu\sigma} l^\sigma &\geq 0, \quad \text{i.e. } T_{\mu\rho} l^\rho \text{ is not space-like.} \end{aligned}$$

  Applied to the perfect fluid this translates to $\rho \geq |p|$ or $\rho = -p$.

- The **strong energy condition**:

$$T_{\mu\nu} t^\mu t^\nu \geq \frac{1}{2} T^\rho_{\ \rho}\, t^\sigma t_\sigma \quad \text{for all time-like vectors } t^\mu.$$

  Applied to the perfect fluid this translates to $\rho + p \geq 0$ and $\rho + 3p \geq 0$.

We may summarise the energy conditions for a perfect fluid as follows: Assuming an equation of state of the form

$$p = w\rho,$$

where w denotes the parameter of the equation of state, and assuming $\rho \geq 0$, each of the energy conditions above implies

$$w \geq -1.$$

9.4 The Robertson-Walker metric

Let us recall the concepts of isotropy and homogeneity of a space: **Isotropy** is the statement that there is no preferred direction in the space, **homogeneity** is the statement that there is no preferred point in the space. Remark: Isotropy and homogeneity are a priori independent concepts, there are manifolds which are homogeneous but nowhere isotropic. An example is the space $\mathbb{R} \times S^2$.

On the other hand we have: If a space is isotropic everywhere, then it is homogeneous. Furthermore we have: If a space is isotropic at one point and in addition homogeneous, then it is isotropic at all points.

From the observation of the cosmic microwave background we may conclude, that the universe as observed from the earth is spatially isotropic at the observation point. As we do not believe that the position of the earth is a preferred point in space, we may assume spatial isotropy of the universe and hence spatial homogeneity of the universe follows.

Remark: We made no implications about the time component. Indeed, we will assume that the universe evolves in time. We will therefore consider a space-time, where the spatial sub-space is homogeneous and isotropic at all times, and the full space-time evolves in time. We may assume (at least locally) that space-time can be written as

$$\mathbb{R} \times \Sigma,$$

where $\mathbb{R}$ represents the time sub-space and Σ a three-dimensional manifold, representing the spatial sub-space. Since the spatial sub-space is homogeneous and isotropic, it follows that Σ must be a maximally symmetric space. By a suitable choice of the time coordinate we may achieve that the metric has the form

$$ds^2 = c^2 dt^2 - R(t)^2 d\sigma^2.$$

$R(t)$ is called the **scale factor**, $d\sigma^2$ denotes the metric on the manifold Σ. We will use the convention that the scale factor $R(t)$ has the dimension of a length, while $d\sigma^2$ is dimensionless. For a maximally symmetric space we have

$$R_{\sigma\lambda\mu\nu} = \kappa \left(g_{\sigma\mu} g_{\lambda\nu} - g_{\sigma\nu} g_{\lambda\mu} \right).$$

We now apply this equation to the three-dimensional space Σ with metric

$$d\sigma^2 = \gamma_{ij} du^i du^j.$$

We find

$$R^{(3)}_{ijkl} = \kappa \left(\gamma_{ik} \gamma_{jl} - \gamma_{il} \gamma_{jk} \right),$$

where the superscript 3 indicates, that we consider the restriction of the curvature tensor to the three-dimensional manifold Σ. The constant κ is given by

$$\kappa = \frac{R^{(3)}}{6}.$$

One obtains for the Ricci tensor

$$Ric^{(3)}_{ij} = 2\kappa\gamma_{ij}.$$

Similar to the case of the Schwarzschild solution, we may put the metric $d\sigma^2$ into the form

$$d\sigma^2 = \gamma_{ij}du^i du^j = e^{2b(r)}dr^2 + r^2 d\Omega^2.$$

As we did for the Schwarzschild solution, we compute from this form the Ricci tensor. In the coordinates (r,θ,φ) we find

$$\begin{aligned} Ric^{(3)}_{11} &= \frac{2}{r}\partial_r b, \\ Ric^{(3)}_{22} &= e^{-2b}\left(r\partial_r b - 1\right) + 1, \\ Ric^{(3)}_{33} &= \left[e^{-2b}\left(r\partial_r b - 1\right) + 1\right]\sin^2\theta. \end{aligned}$$

Equating the above equations to $Ric^{(3)}_{ij} = 2\kappa\gamma_{ij}$, we may solve for $b(r)$. We first consider $Ric^{(3)}_{11}$:

$$\begin{aligned} \frac{2}{r}\partial_r b &= 2\kappa e^{2b}, \\ e^{-2b}db &= \kappa r dr, \\ -\frac{1}{2}e^{-2b} &= \frac{1}{2}\kappa r^2 - \frac{1}{2}c_0, \\ b(r) &= -\frac{1}{2}\ln\left(c_0 - \kappa r\right), \end{aligned}$$

with some yet unknown integration constant c_0. In order to fix c_0, we consider $Ric^{(3)}_{22}$:

$$e^{-2b}\left(r\partial_r b - 1\right) + 1 = \left(c_0 - \kappa r\right)\left[\frac{\kappa r^2}{\left(c_0 - \kappa r\right)} - 1\right] + 1 = 2\kappa r^2 - c_0 + 1.$$

This should be equal to $2\kappa r^2$ and it follows that $c_0 = 1$. We therefore have

$$b(r) = -\frac{1}{2}\ln\left(1 - \kappa r^2\right)$$

and hence

$$d\sigma^2 = \frac{dr^2}{1 - \kappa r^2} + r^2 d\Omega^2.$$

Combining all results, we obtain for the metric of four-dimensional space-time

$$ds^2 = c^2 dt^2 - R(t)^2 \left[\frac{dr^2}{1-\kappa r^2} + r^2 \left(d\theta^2 + \sin^2\theta d\phi^2 \right) \right].$$

A metric of this form is called **Robertson-Walker metric**. We recall that we use the convention that $R(t)$ has dimension of a length. Then the coordinates (r, θ, ϕ) and the parameter κ are dimensionless. In particular r is dimensionless. The Robertson-Walker metric is invariant under a rescaling

$$R \to \lambda^{-1} R, \qquad r \to \lambda r, \qquad \kappa \to \lambda^{-2}\kappa.$$

We may use this rescaling to convert to the convention, where r has the dimension of a length, R is dimensionless and κ has the dimension length^{-2}.

If we stick to our original convention, where r and κ are dimensionless, we may use the rescaling to rescale κ to $\{-1, 0, 1\}$.

We still have to determine the scale factor $R(t)$. The possible geometries can be divided into three classes according to the parameter κ:

$\kappa = 1$	(or more general $\kappa > 0$)	closed geometry
$\kappa = 0$		spatially flat
$\kappa = -1$	(or more general $\kappa < 0$)	open geometry

This is most easily seen by introducing a new radial coordinate through

$$d\chi = \frac{dr}{\sqrt{1-\kappa r^2}}.$$

Upon integration of this equation we obtain

$$r = \begin{cases} \sin\chi, & \kappa = 1, \\ \chi, & \kappa = 0, \\ \sinh\chi, & \kappa = -1. \end{cases}$$

and hence

$$d\sigma^2 = \begin{cases} d\chi^2 + \sin^2\chi d\Omega^2, & \kappa = 1, \\ d\chi^2 + \chi^2 d\Omega^2, & \kappa = 0, \\ d\chi^2 + \sinh^2\chi d\Omega^2, & \kappa = -1. \end{cases}$$

For $\kappa = 1$ we obtain for $d\sigma^2$ the metric of the sphere S^3, for $\kappa = 0$ we obtain the flat Euclidean metric and for $\kappa = -1$ we obtain a hyperbolic metric.

9.5 Friedmann equations and the Hubble parameter

In order to determine the function $R(t)$ we now use Einstein's equations and the model of a perfect fluid for the energy-momentum tensor

$$T_{\mu\nu} = (p+\rho)\, u_\mu u_\nu - p g_{\mu\nu}.$$

together with the equation of state

$$p = w\rho.$$

In the rest frame of the fluid we have $u^\mu = (1,0,0,0)$. We use the Robertson-Walker metric to lower the index and we obtain (as in flat space) $u_\mu = (1,0,0,0)$. Hence

$$T_{\mu\nu} = \begin{pmatrix} \rho & 0 & 0 & 0 \\ 0 & & & \\ 0 & & -p g_{ij} & \\ 0 & & & \end{pmatrix}$$

For the trace we have

$$T = T^\mu_\mu = g^{\mu\nu} T_{\mu\nu} = (p+\rho)\, u^\mu u_\mu - p g^{\mu\nu} g_{\mu\nu} = (p+\rho) - 4p = \rho - 3p.$$

We recall that Einstein's equations may be written as

$$Ric_{\mu\nu} = \frac{8\pi G}{c^4}\left(T_{\mu\nu} - \frac{1}{2} g_{\mu\nu} T\right).$$

For the $(\mu,\nu) = (0,0)$-component one finds

$$-3\frac{1}{c^2}\frac{\ddot{R}}{R} = \frac{4\pi G}{c^4}(\rho + 3p)$$

and for the $(\mu,\nu) = (i,j)$-components one obtains

$$\frac{1}{c^2}\frac{\ddot{R}}{R} + 2\frac{1}{c^2}\left(\frac{\dot{R}}{R}\right)^2 + 2\frac{\kappa}{R^2} = \frac{4\pi G}{c^4}(\rho - p).$$

(A dot over a function denotes the time derivative d/dt.) We may eliminate the second time derivative from the last equation. This gives us **Friedmann's equations**:

$$\left(\frac{\dot{R}}{R}\right)^2 = \frac{8\pi G\rho}{3c^2} - \frac{\kappa c^2}{R^2},$$
$$\frac{\ddot{R}}{R} = -\frac{4\pi G}{3c^2}(\rho + 3p).$$

We call the quantity

$$H(t) = \frac{\dot{R}(t)}{R(t)}$$

the **Hubble parameter**. The value of the Hubble parameter at our current time is called the Hubble constant H_0. The value of the Hubble constant is

$$H_0 = 67.8 \pm 0.9\,\mathrm{km\,s^{-1}Mpc^{-1}} \qquad \text{(Planck satellite).}$$

One Megaparsec equals $1\mathrm{Mpc} = 3.09 \cdot 10^{22}\mathrm{m}$. The Hubble parameter is a measure for the expansion of the universe.

The rate, by which the expansion of the universe slows down, is described by the parameter

$$q = -\frac{R\ddot{R}}{\dot{R}^2} = -\frac{\frac{\ddot{R}}{R}}{\left(\frac{\dot{R}}{R}\right)^2}.$$

From Friedmann's equations we have

$$q = \frac{4\pi G}{3c^2}\frac{(\rho + 3p)}{H(t)^2} = \frac{\rho + 3p}{2\rho - \frac{3c^4\kappa}{4\pi G R^2}}.$$

For the time variation of the Hubble parameter we have

$$\dot{H}(t) = \frac{d}{dt}\frac{\dot{R}(t)}{R(t)} = \frac{\ddot{R}}{R} - \left(\frac{\dot{R}}{R}\right)^2 = -\left[1 + q(t)\right] H(t)^2.$$

In terms of the Hubble parameter Friedmann's equations read

$$\begin{aligned} H(t)^2 &= \frac{8\pi G\rho}{3c^2} - \frac{\kappa c^2}{R^2}, \\ \dot{H}(t) &= -\left[1 + q(t)\right] H(t)^2. \end{aligned}$$

The **critical density** is defined by $\kappa = 0$, hence

$$\rho_c = \frac{3c^2 H^2}{8\pi G}$$

As the Hubble parameter is time-dependent, so is the critical density. We also define a **density parameter** Ω by

$$\Omega = \frac{\rho}{\rho_c} = \frac{8\pi G}{3c^2 H^2}\rho.$$

With these definitions the first Friedmann equation may be written as

$$\Omega = 1 + \kappa \frac{c^2}{H^2 R^2}.$$

We therefore have:

$$\begin{array}{ccccccc}
\rho < \rho_c & \leftrightarrow & \Omega < 1 & \leftrightarrow & \kappa < 0 & \leftrightarrow & \text{open}, \\
\rho = \rho_c & \leftrightarrow & \Omega = 1 & \leftrightarrow & \kappa = 0 & \leftrightarrow & \text{flat}, \\
\rho > \rho_c & \leftrightarrow & \Omega > 1 & \leftrightarrow & \kappa > 0 & \leftrightarrow & \text{closed}.
\end{array}$$

9.6 Evolution of the universe

We start from energy conservation, or more concretely from

$$\nabla_\mu T^\mu_\nu = 0.$$

For the $\nu = 0$ component we have

$$0 = \partial_\mu T^\mu_{\ 0} + \Gamma^\mu_{\mu\lambda} T^\lambda_{\ 0} - \Gamma^\lambda_{\mu 0} T^\mu_{\ \lambda} = \partial_0 \rho - 3 \frac{\dot{R}}{R} (\rho + p).$$

With the equation of state $p = w\rho$ we obtain

$$\frac{\dot{\rho}}{\rho} = -3(1+w) \frac{\dot{R}}{R},$$

or

$$\frac{d \ln \rho(t)}{dt} = -3(1+w) \frac{d \ln R(t)}{dt}.$$

We may integrate this equation and obtain

$$\rho(t) = \rho(t_0) \left(\frac{R(t)}{R(t_0)} \right)^{-3(1+w)}.$$

We consider a few special cases: Let us first assume that the universe consists of non-interacting galaxies (dust). We have $w = 0$ and

$$\rho_M(t) \sim R(t)^{-3}.$$

A universe where the energy density decreases as $R(t)^{-3}$ is called a **matter dominated universe**. As a second example let us consider a universe consisting solely of photons. We now have $w = 1/3$ and

$$\rho_R(t) \sim R(t)^{-4}.$$

A universe where the energy density decreases as $R(t)^{-4}$ is called a **radiation dominated universe**.
As a final example we consider a universe, which consists solely of vacuum energy. In this case we have $w = -1$ and

$$\rho_\Lambda(t) \quad \sim \quad R(t)^0.$$

A universe where the energy density is constant as a function of time is called a **vacuum dominated universe**.

In all cases we find a power law

$$\rho(t) \quad = \quad \rho_0 \left(\frac{R(t)}{R_0}\right)^{-n},$$

with $n = 3(1+w)$. Let us return to the first Friedmann equation:

$$H(t)^2 \quad = \quad \frac{8\pi G}{3c^2}\rho(t) - \frac{\kappa c^2}{R(t)^2}$$

We may interpret the term proportional to the spatial curvature κ as an effective energy density

$$\rho_{curv}(t) \quad = \quad -\frac{3c^4\kappa}{8\pi G}R(t)^{-2}.$$

With $n = 3(1+w)$ we have in this case $w = -1/3$. We further set

$$\Omega_{curv} \quad = \quad \frac{\rho_{curv}}{\rho_c} = -\frac{c^2\kappa}{R(t)^2 H(t)^2} = -\frac{c^2\kappa}{\dot{R}^2}.$$

With these conventions we have

$$H(t)^2 \quad = \quad \frac{8\pi G}{3c^2}\left(\rho_{curv}(t) + \sum_j \rho_j(t)\right).$$

Dividing both sides by $H(t)^2$ one obtains

$$1 \quad = \quad \Omega_{curv} + \sum_j \Omega_j.$$

Remark: The total energy density of the universe is of course just

$$\Omega \quad = \quad \sum_j \Omega_j,$$

i.e. without Ω_{curv}. We therefore have

$$\Omega_{curv} \quad = \quad 1 - \Omega.$$

The introduction of ρ_{curv} and Ω_{curv} only serves to unify the discussion of the various contributions to $H(t)$.
Let us now consider for simplicity an energy density with the time-dependence

$$\rho(t) \;=\; \rho_0 \left(\frac{R(t)}{R_0}\right)^{-n}.$$

For $n > 0$ we obtain from

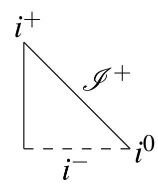
$$H(t)^2 \;=\; \frac{8\pi G}{3c^2}\rho(t)$$

$$\dot{R}(t) \;=\; \sqrt{\frac{8\pi G}{3c^2}\rho_0 R_0^n}\; R(t)^{1-\frac{n}{2}}$$

and hence

$$R(t) \;=\; R_0 \left[\frac{2}{3c^2} n^2 \pi G \rho_0 (t-t_0)^2\right]^{\frac{1}{n}}.$$

At time $t = t_0$ we have $R(t_0) = 0$. This space-time has a true singularity at $t = t_0$. This can be seen by considering for example the energy density $\rho(t) \sim R(t)^{-n}$, which diverges for $n > 0$ at $t = t_0$. The singularity at $t = t_0$ is called the **big bang**. The associated Penrose diagram is given by



Finally, let us consider the special case $n = 0$, corresponding to a universe consisting of vacuum energy. In this case the energy density is constant

$$\rho(t) \;=\; \rho_0$$

and we find

$$R(t) \;=\; R_0 \exp\left(\sqrt{\frac{8}{3c^2}\pi G \rho_0}\,(t-t_0)\right).$$

Let us summarise the essential features of the various contributions to the right-hand side of the first Friedmann equation:

$$\begin{array}{lllll}
\text{radiation :} & n=4, & w=\frac{1}{3}, & \rho \sim R^{-4}, & R \sim t^{\frac{1}{2}}, \\
\text{matter :} & n=3, & w=0, & \rho \sim R^{-3}, & R \sim t^{\frac{2}{3}}, \\
\text{curvature :} & n=2, & w=-\frac{1}{3}, & \rho \sim R^{-2}, & R \sim t, \\
\text{vacuum energy :} & n=0, & w=-1, & \rho \sim R^{0}, & R \sim e^{t}.
\end{array}$$

The relevant equations are $p = w\rho$, $n = 3(1+w)$, $\rho \sim R^{-n}$ and (for $n > 0$) $R \sim t^{2/n}$.

9.7 The red shift

For simplicity we consider a spatially flat universe ($\kappa = 0$) with the Robertson-Walker metric

$$ds^2 = c^2 dt^2 - R(t)^2 \left[dr^2 + r^2 \left(d\theta^2 + \sin^2\theta d\phi^2 \right) \right].$$

Let us for the moment further assume that the time-dependence of the scale factor is given by a power law:

$$R(t) = R_0 \left(\frac{t}{t_0} \right)^q, \quad 0 < q < 1,$$

where R_0 is a quantity with the dimension of a length and t_0 is a quantity with the dimension of a time interval. The power q is given for a perfect fluid with equation of state $p = w\rho$ by

$$q = \frac{2}{n} = \frac{2}{3(1+w)}.$$

For example, for a radiation dominated universe we have $q = 1/2$, while for a matter dominated universe we have $q = 2/3$. Light cones in a curved space-time are defined by null paths, i.e. $ds^2 = 0$. For the light propagation we obtain

$$\frac{dx}{dt} = \pm c \left(\frac{t}{t_0} \right)^{-q}.$$

Here we have introduced $x = R_0 r$. Recall that with our convention r is a dimensionless quantity. x has dimension of a length. The equation above can be integrated and one obtains

$$t = \left[\frac{(1-q)}{c\, t_0^q} (\pm x - x_0) \right]^{\frac{1}{1-q}}.$$

Let us discuss the most important properties of this solution: The light cones at $t = 0$ are tangential to the singularity at $t = 0$:

$$\frac{dt}{dx} = \pm \frac{1}{c} \left(\frac{t}{t_0} \right)^q = 0 \quad \text{for } t = 0 \text{ and } q > 0.$$

A second important property of this geometry is given by the fact, that the past light cones of two distinct points are not required to intersect. If there is no intersection, the two points are not in **causal contact**. This is in contrast to flat Minkowski space, where the past light cones always intersect.

Let us now consider light propagation in curved space-time without assuming a power law for the scale factor. We will however again assume a spatially flat space-time. We start from the Robertson-Walker metric

$$ds^2 = c^2 dt^2 - R(t)^2 \left[dr^2 + r^2 \left(d\theta^2 + \sin^2\theta d\phi^2 \right) \right]$$

and consider the geodesic equation

$$\frac{d^2x^\mu}{d\lambda^2} + \Gamma^\mu_{\tau\sigma}\frac{dx^\tau}{d\lambda}\frac{dx^\sigma}{d\lambda} = 0$$

For a massless particle (photon) we have

$$\frac{dx^\mu}{d\lambda}\frac{dx_\mu}{d\lambda} = 0.$$

It will be convenient to normalise the curve parameter λ such that

$$p^\mu = \frac{dx^\mu}{d\lambda}.$$

An observer with four-velocity u^μ measures a photon energy given by

$$E = cp_\mu u^\mu.$$

We obtain for the 0-component of the geodesic equation

$$c\frac{d^2t}{d\lambda^2} + \frac{1}{c}R\dot{R}\left(\frac{dr}{d\lambda}\right)^2 = 0.$$

With

$$\frac{dr}{d\lambda} = \frac{c}{R}\frac{dt}{d\lambda}$$

one obtains

$$\frac{d^2t}{d\lambda^2} + \frac{\dot{R}}{R}\left(\frac{dt}{d\lambda}\right)^2 = 0.$$

A solution is given by

$$\frac{dt}{d\lambda} = \frac{c_0}{R(t)},$$

where c_0 is a constant. Let us verify this solution:

$$\begin{aligned}
\frac{d^2t}{d\lambda^2} + \frac{\dot{R}}{R}\left(\frac{dt}{d\lambda}\right)^2 &= \frac{d}{d\lambda}\frac{c_0}{R} + \frac{\dot{R}}{R}\frac{c_0^2}{R^2} = \left(\frac{d}{dt}\frac{c_0}{R}\right)\frac{dt}{d\lambda} + \frac{c_0^2\dot{R}}{R^3} \\
&= \left(\frac{-c_0\dot{R}}{R^2}\right)\frac{c_0}{R} + \frac{c_0^2\dot{R}}{R^3} = 0.
\end{aligned}$$

We will see in a second that $c_0 = E_0 R(t_0)/c^2$, where E_0 is the energy of the photon at some initial time t_0 and $R(t_0)$ is the scale factor at this time. An observer with constant spatial coordinates (and hence four-velocity $u^\mu = (1,0,0,0)$) measures a photon with energy

$$E = cu^\mu\frac{dx_\mu}{d\lambda} = c^2\frac{c_0}{R(t)}.$$

This implies the **cosmological red shift**: A photon emitted with energy E_1 at a time t_1 with scale factor $R(t_1)$ and measured with energy E_2 at a time t_2 with scale factor $R(t_2)$ satisfies the relation

$$\frac{E_2}{E_1} = \frac{R(t_1)}{R(t_2)}.$$

The name "red shift" derives from the fact that in an expanding universe we have $R(t_2) > R(t_1)$ for $t_2 > t_1$. This implies $E_2 < E_1$. Usually the red shift is denoted as

$$z = \frac{E_1 - E_2}{E_2} = \frac{\lambda_2 - \lambda_1}{\lambda_1} = \frac{R(t_2)}{R(t_1)} - 1.$$

Thus

$$\frac{R(t_1)}{R(t_2)} = \frac{1}{1+z}.$$

Given the red shift and the scale factor at the time of the observation, we may deduce the scale factor at the time of the emission of the photon.

Remark: The red shift and the Doppler effect are conceptually different: The Doppler effect requires a flat space, such that the relative velocity between two objects is well defined. On a curved manifold, we may only compare tangent vectors at the same space-time point, a relative velocity between two distant points is not well-defined. The cosmological red shift is entirely due to the change in the metric.

With this warning, we are now nevertheless going to associate a velocity to the red shift. We first introduce the **instantaneous physical distance** $d_p(t)$ between two objects (e.g. galaxies). If the first object is located at the spatial origin and the second object has the radial coordinate r, we define

$$d_p(t) = R(t) r.$$

The rate of change of the instantaneous physical distance defines a velocity

$$v = \dot{d}_p(t) = \dot{R}(t) r = \frac{\dot{R}(t)}{R(t)} d_p(t) = H(t) d_p(t).$$

This is **Hubble's law**.

The instantaneous physical distance is not an observable, as observations always refer to our past light-cone. In practice, the **luminosity distance** d_L is used. Suppose we know the luminosity of some object (e.g. stars or galaxies) and measure the photon flux, then we can infer the distance. In an Euclidean space we have

$$d_L^2 = \frac{L}{4\pi F},$$

where L is the luminosity of the source (i.e. emitted energy per unit time) and F the observed flux (i.e. energy per unit time per unit area). The formula just says that the energy emitted by the source per unit time is the same as the energy through a sphere with radius d_L per unit time.

Let us now adapt this formula to the Robertson-Walker metric. It will be convenient to use χ instead of r as radial variable. The metric reads

$$\begin{aligned} ds^2 &= c^2 dt^2 - R(t)^2 \left[\frac{dr^2}{1-\kappa r^2} + r^2 \left(d\theta^2 + \sin^2\theta d\phi^2 \right) \right] \\ &= c^2 dt^2 - R(t)^2 \left[d\chi^2 + S_\kappa(\chi)^2 \left(d\theta^2 + \sin^2\theta d\phi^2 \right) \right], \end{aligned}$$

with

$$S_\kappa(\chi) = \begin{cases} \sin\chi, & \kappa = 1, \\ \chi, & \kappa = 0, \\ \sinh\chi, & \kappa = -1. \end{cases}$$

The relation between r and χ is $r = S_\kappa(\chi)$.

Conservation of the photon number tells us, that all photons emitted from the source will eventually pass through the sphere with radius χ. However, we have to take two effects into account: First of all, photons emitted with an energy E are red-shifted to the energy $E/(1+z)$. Secondly, the photons arrive less frequently at the sphere: Photons emitted a time Δt apart will hit the sphere a time $(1+z)\Delta t$ apart. Thus

$$L = (1+z)^2 AF,$$

where A is the area of the sphere, given by

$$A = 4\pi R_0^2 S_\kappa(\chi)^2.$$

R_0 is the scale factor at the observation time t_0. Thus

$$d_L = (1+z) R_0 S_\kappa(\chi).$$

The radial variable χ is not an observable and we would like to eliminate this variable in favour of measurable quantities. We can do this as follows: Consider a radial null geodesic:

$$0 = ds^2 = c^2 dt^2 - R(t)^2 d\chi^2.$$

We have

$$\frac{d\chi}{dt} = \frac{c}{R(t)}$$

and therefore

$$\chi = c \int\limits_t^{t_0} \frac{dt'}{R(t')} = c \int\limits_R^{R_0} \frac{dR'}{R'^2 H(R')} = \frac{c}{R_0} \int\limits_0^z \frac{dz'}{H(z')},$$

where we first changed variables from t to the scale factor R, and then from the scale factor R to the red shift z with $R = R_0/(1+z)$. Let us define

$$E(z) = \frac{H(z)}{H_0}.$$

$E(z)$ is a dimensionless quantity and given by

$$E(z) = \frac{1}{H_0}\left(\frac{8\pi G}{3c^2}\sum_i \rho_i(z)\right)^{\frac{1}{2}}.$$

The sum over i includes the curvature component. If all energy densities evolve with power laws

$$\rho_i(z) = \rho_{i,0}\left(\frac{R(z)}{R_0}\right)^{-n_i} = \rho_{i,0}(1+z)^{n_i},$$

we have

$$E(z) = \frac{1}{H_0}\left(\frac{8\pi G}{3c^2}\sum_i \rho_{i,0}(1+z)^{n_i}\right)^{\frac{1}{2}} = \left(\sum_i \Omega_{i,0}(1+z)^{n_i}\right)^{\frac{1}{2}}$$

Putting everything together we obtain

$$d_L = (1+z)R_0 S_\kappa(\chi) = (1+z)R_0 S_\kappa\left(\frac{c}{R_0 H_0}\int\limits_0^z \frac{dz'}{E(z')}\right).$$

If $\kappa = 0$ we have $S_\kappa(\chi) = \chi$ and R_0 drops out. We obtain in this case

$$d_L = \frac{c(1+z)}{H_0}\int\limits_0^z \frac{dz'}{E(z')}.$$

For $\kappa \in \{1,-1\}$ we eliminate R_0 in favour of $\Omega_{curv,0}$:

$$\Omega_{curv,0} = -\frac{c^2\kappa}{R_0^2 H_0^2} \Rightarrow R_0 = \frac{c}{H_0\sqrt{|\Omega_{curv,0}|}}.$$

This gives

$$d_L = \frac{c(1+z)}{H_0\sqrt{|\Omega_{curv,0}|}} S_\kappa\left(\sqrt{|\Omega_{curv,0}|}\int\limits_0^z \frac{dz'}{E(z')}\right).$$

This formula is of central importance in cosmology. Given H_0 and $\Omega_{i,0}$ we may calculate the luminosity distance as a function of z. If we measure both the red shift z and the luminosity distance for a number of objects, we may extract information on H_0 and $\Omega_{i,0}$.

We may also ask at what time t a photon was emitted, which is observed today at time t_0 with red shift z. t is called the **look-back time**. We have

$$t_0 - t = \int\limits_t^{t_0} dt' = \int\limits_R^{R_0} \frac{dR'}{R' H(R')} = \int\limits_0^z \frac{dz'}{(1+z') H(z')} = \frac{1}{H_0} \int\limits_0^z \frac{dz'}{(1+z') E(z')}$$

As a simple example consider a flat matter-dominated universe. Then

$$E(z) = (1+z)^{\frac{3}{2}}$$

and

$$t_0 - t = \frac{1}{H_0} \int\limits_0^z dz' \left(1+z'\right)^{-\frac{5}{2}} = -\frac{2}{3H_0} \left[(1+z)^{-\frac{3}{2}} - 1\right]$$

In the limit $z \to \infty$ we find

$$\lim_{z\to\infty} (t_0 - t) = \frac{2}{3H_0},$$

which gives the total age of a flat matter-dominated universe. A value of $H_0 = 70\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ gives the age $9.3 \cdot 10^9$ yr, which is not too far off from the actual age $13.8 \cdot 10^9$ yr.

Let us define the **particle horizon** and the **event horizon**: Using the coordinates (ct, χ, θ, ϕ) we consider an observer and an emitter

$$\begin{aligned} \text{observer} &: \ \chi = 0,\ \theta = 0,\ \phi = 0, \\ \text{emitter} &: \ \chi = \chi_p,\ \theta = 0,\ \phi = 0. \end{aligned}$$

We may ask, what is the value of the radial variable χ_p, such that signals emitted by the emitter at an initial time t_i (very often we will take $t_i = 0$) can no longer reach the observer at time t_0 (usually the time of today). This defines the **particle horizon**. χ_p is given by

$$\chi_p = c \int\limits_{t_i}^{t_0} \frac{dt}{R(t)} = c \int\limits_{R_i}^{R_0} \frac{dR}{R^2 H(R)}.$$

χ_p is a dimensionless quantity. In order to get the value of the particle horizon today in units of length, one multiplies with R_0:

$$d_p = R_0 \chi_p.$$

If we take $t_i = 0$ this means: Points, which are more than d_p away from us today, cannot have affected us up to today (however they may affect us in the future).
If the scale factor grows as $R \sim t^{2/n}$ ($n = 3$ for matter, $n = 4$ for photons), the particle horizon grows as

$$\chi_p \sim t^{1-\frac{2}{n}}.$$

The second question which we may ask is the following: What is the value of the radial variable χ_e, such that signals emitted by the emitter at time t_0 (usually the time of today) will not reach us until a final time t_F (very often we will take $t_f = \infty$). This defines the **event horizon**. χ_e is given by

$$\chi_e = c \int\limits_{t_0}^{t_f} \frac{dt}{R(t)} = c \int\limits_{R_0}^{R_f} \frac{dR}{R^2 H(R)}.$$

In order to get the value of the event horizon today in units of length, one multiplies with R_0:

$$d_e = R_0 \chi_e.$$

If we take $t_f = \infty$ this means: Points, which are more than d_e away from us today, cannot affect us in the future.

9.8 The cosmic microwave background

Consider a universe consisting of photons, electrons and protons. The photons scatter off the charged electrons and protons through Thomson scattering. They are in thermal equilibrium and their spectral energy density (i.e. energy per volume and per unit frequency) is described by Planck's law for black-body radiation:

$$u(\omega, T) = \frac{\hbar}{\pi^2 c^3} \frac{\omega^3}{e^{\frac{\hbar\omega}{k_B T}} - 1}.$$

Integration over the frequency ω gives the Stefan-Boltzmann law:

$$\rho = \frac{4\sigma}{c} T^4,$$

with the Stefan-Boltzmann constant

$$\sigma = \frac{\pi^2 k_B^4}{60 \hbar^3 c^2}.$$

The universe appears opaque with respect to electromagnetic radiation: Photons do not propagate freely, but scatter frequently. Assume now that the universe is cooling down and electrons and protons combine to (neutral) hydrogen atoms. This is called the **recombination epoch**. Being neutral, the (low-energy) photons do not scatter on the hydrogen atoms. After sufficient many charged particles combine to form neutral particles, the photons **decouple**: There are simply no charged interaction partners left. The photons now propagate freely and the universe becomes transparent to electromagnetic radiation. As the universe further expands, the photons are red-shifted to lower energies and observed as the cosmic microwave background.

Given an initial spectral energy density $u(\omega_1, T_1)$ at time t_1 (at decoupling time) we would like to derive the spectral energy density $u(\omega_2, T_2)$ at time t_2 (today). Consider a photon with energy

$$E_1 = \hbar\omega$$

at decoupling. With a red shift z the observed energy (or frequency) today is

$$E_2 = \frac{E_1}{1+z}, \qquad \omega_2 = \frac{\omega_1}{1+z}.$$

As the photons do not interact, the number of photons stays constant. However, the universe expands. A comoving volume changes from V_1 to

$$V_2 = \left(\frac{R(t_2)}{R(t_1)}\right)^3 V_1 = (1+z)^3 V_1.$$

Combining everything we find (please note that "energy per unit frequency" is invariant under a simultaneous rescaling of the energy and the frequency)

$$\begin{aligned} u(\omega_2, T_2) &= (1+z)^{-3} u(\omega_1, T_1) = (1+z)^{-3} u((1+z)\omega_2, T_1) = \frac{\hbar}{\pi^2 c^3} \frac{\omega_2^3}{e^{\frac{(1+z)\hbar\omega_2}{k_B T_1}} - 1} \\ &= u\left(\omega_2, \frac{T_1}{1+z}\right). \end{aligned}$$

Thus the spectrum of the black-body radiation is conserved, however the corresponding temperature is lowered:

$$T_2 = \frac{T_1}{1+z}.$$

The temperature T_2 is very well measured ($T_2 = 2.73\mathrm{K}$). In addition, the typical energy scale where decoupling occurs is known: ~ 1 eV. We may therefore deduce the red shift and the ratio of the scale parameters $R(t_1)/R(t_2)$.

To a first approximation the observed cosmic microwave background is isotropic. However, precise measurements reveal an anisotropy at the order of 10^{-5}.

9.9 The current paradigm for our universe

The measured density parameters are

$$\begin{aligned} \Omega_\gamma &= (5.4 \pm 0.1) \cdot 10^{-5}, \\ \Omega_M &= 0.31 \pm 0.01, \\ \Omega_{vac} &= 0.69 \pm 0.01. \end{aligned}$$

The matter density parameter is the sum of the baryonic matter density parameter (ordinary matter) and the dark matter density parameter

$$\Omega_M = \Omega_B + \Omega_{DM},$$

with the values

$$\Omega_B = 0.048 \pm 0.001,$$
$$\Omega_{DM} = 0.26 \pm 0.01.$$

This gives a value of

$$\Omega_{curv} = 1 - \Omega_{vac} - \Omega_M - \Omega_\gamma$$

compatible with $\kappa = 0$, i.e. a spatially flat universe.

The age of the universe is

$$\tau = (13.80 \pm 0.04) \cdot 10^9 \text{yr}.$$

The **large scale structure of the universe**: Stars assemble in galaxies, galaxies form clusters and clusters form super-clusters.

Cornerstones of the universe:

Event	Time	Energy	Temperature	Red shift
Big bang	0 s			
Planck era	$< 10^{-43}$ s	$> 10^{18}$ GeV	$> 10^{31}$ K	
Inflation	$\gtrsim 10^{-34}$ s	$\lesssim 10^{15}$ GeV	$\lesssim 10^{28}$ K	
Baryogenesis	$< 10^{-10}$ s	> 1 TeV	$> 10^{16}$ K	
Electroweak symmetry breaking	10^{-10} s	1 TeV	10^{16} K	
Quark-hadron transition	10^{-4} s	100 MeV	10^{12} K	
Nucleon freeze-out	10^{-2} s	10 MeV	10^{11} K	
Neutrino decoupling	1 s	1 MeV	10^{10} K	
Big bang nucleosynthesis	3 min	100 keV	10^9 K	
Matter-radiation equality	10^4 yr	1 eV	10^4 K	10^4
Recombination	10^5 yr	0.3 eV	$3 \cdot 10^3$ K	1100
Dark Ages	$10^5 - 10^8$ yr	> 6 meV	> 70 K	> 25
Reionisation	10^8 yr	$1.5 - 6$ meV	$20 - 70$ K	$6 - 25$
Galaxy formation	$\sim 6 \cdot 10^8$ yr	~ 2.6 meV	~ 30 K	~ 10
Dark energy dominates	$\sim 10^9$ yr	~ 0.7 meV	~ 8 K	~ 2
Solar system	$8 \cdot 10^9$ yr	0.35 meV	4 K	0.5
Today	$14 \cdot 10^9$ yr	0.24 meV	2.73 K	0

Please note that energy, temperature and red shift are related: If E_0 and T_0 denote the energy and the temperature of the universe today, the corresponding values at red shift z are given by

$$E = (1+z)E_0, \qquad T = (1+z)T_0.$$

The relation between E and T is $E = k_B T$.

During the "dark ages" epoch, galaxies and stars gradually form through gravitational interactions. As there are no visible stars yet at the beginning of this era, the epoch is called "dark ages". At the end of this epoch, high energy photons from the first stars can ionise hydrogen in the inter-galactic medium. This is called "reionisation".

10 Dark matter and thermal relics

10.1 Basic facts about dark matter

1. Dark matter has **attractive gravitational interactions**, hence the name “matter”. Evidence for dark matter is provided by:

   - On galactic scales: Observation of flat rotation curves of disk galaxies.
   - On cluster scales: Observation of the velocity dispersion of galaxies in the Coma Cluster.
   - On cosmological scales: Measurement of the matter density parameter: $\Omega_M = \Omega_B + \Omega_{DM}$ with $\Omega_B \approx 0.05$ and $\Omega_{DM} \approx 0.26$.

   There is no evidence that dark matter has any other interaction but gravity.

2. Dark matter is either **stable** or has a lifetime larger than the age of the universe. Otherwise it wouldn’t be here today.

3. Dark matter is **not observed to interact with light**, hence the name “dark”. This implies that the coupling to the electromagnetic field is either small and/or the dark matter particles are heavy.

4. The major part of dark matter must be **dissipationless**. “Dissipationless” means that dark matter particles cannot cool down by emitting particles like photons. If dark matter would be dissipative, the dark halos would not exist.

   Galaxy formation starts from a mixture of ordinary and dark matter. The visible matter dissipates energy by emitting photons and falls into the potential well of the object. Because the emission is isotropic, the angular momentum of the visible matter is preserved. Thus as the visible matter collapses to the centre, it increases its angular speed until it becomes unstable towards the formation of a disk, which thus rotates much faster than the dark halo.

5. The **mass** m of the major component of dark matter is bounded by

   $$m \quad < \quad 2 \cdot 10^{48}\,\text{GeV}.$$

   This is a very weak constraint. This bound comes from the non-observation of massive astrophysical compact halo objects (MACHOS). in the dark halo of our galaxy.

6. Dark matter is usually assumed to be **collisionless**, however the limit on dark matter self-interactions is very large:

   $$\frac{\sigma_{\text{self}}}{m} \quad \leq \quad 2\,\text{barn GeV}^{-1}.$$

   The limit comes from two colliding galaxies in the bullet cluster.

7. The bulk of dark matter is either **cold** or warm. Dark matter is classified as hot, warm or cold according to how relativistic it was when the temperature of the universe was of the order of $\simeq$ keV. Hot dark matter is relativistic at that time, cold dark matter is non-relativistic at that time and warm dark matter just turns from relativistic to non-relativistic at that time. Simulations of the formation of the large scale structure of our universe shows that cold dark matter models are compatible with the observed large scale structure, while hot dark matter models are not.

   Baryonic matter can only cluster after recombination, before recombination the photon pressure in the plasma prevents it. However, shortly after recombination baryonic matter must be attracted by already existing inhomogeneities of dark matter, otherwise there would be not enough time to form the structures we observe now.

   Stars and galaxies should form first, while clusters and super-clusters should form second. This requires galaxy-size dark matter inhomogeneities to survive the horizon crossing (i.e. when $\chi_{\mathrm{galaxy}} = \chi_p$, which corresponds to the temperature being $\simeq$ keV). After horizon crossing, the inhomogeneities could potentially be washed out. This happens for hot dark matter. However this does not happen, if dark matter is cold or warm.

   Simulations of hot dark matter show, that in these models super-clusters and clusters form fist and later fragment into galaxies.

8. Most dark matter candidates are **relics from pre-big bang nucleosynthesis**. This implies that the calculation of the dark matter relic abundance or the primordial dark matter velocity distribution depends on assumptions on the thermal history of the universe. With different viable assumptions, the relic density and velocity distribution may change considerably.

10.2 Thermal freeze-out

Let us discuss a dark matter particle X together with its anti-particle $\bar{X}$. The dark matter particle and the anti-particle may annihilate, let us assume that the reaction is

$$X + \bar{X} \quad \rightarrow \quad Y + \bar{Y},$$

where Y and $\bar{Y}$ are two Standard Model particles. The inverse reaction is the production process

$$Y + \bar{Y} \quad \rightarrow \quad X + \bar{X}.$$

We say that the particles are in **chemical equilibrium**, if the production and annihilation processes occur at the same rate, i.e. on the average the particle numbers are conserved. Let us also consider an elastic scattering process like

$$X + Y \quad \rightarrow \quad X + Y,$$

If elastic scattering processes occur frequently enough, the particles are in **kinetic equilibrium**. Please note that it is possible that particles are no longer in chemical equilibrium, but maintain kinetic equilibrium.

If $m_X \gg m_Y$ we would expect that at low temperatures most dark matter particles would have annihilated into Standard Model particles. We have to find a mechanism, which explains the dark matter energy density.

Let us first discuss the mechanisms for baryons and photons:

1. **Baryon-anti-baryon asymmetry**: It is generally believed, that initially there has been roughly the same number of baryons as anti-baryons with a tiny asymmetry, making the number of baryons slightly higher than the number of anti-baryons, i.e. at times $t \leq 10^{-6}$s

$$\frac{n_q - n_{\bar{q}}}{n_q} \simeq 3 \cdot 10^{-8}.$$

   All anti-baryons annihilate with a baryon, such that the tiny surplus of baryons survives and constitutes the matter we observe today.

2. **Photon decoupling**: Before photon decoupling, photons are in thermal equilibrium through elastic scattering processes like

$$\gamma + e^- \rightarrow \gamma + e^- \quad \text{or} \quad \gamma + p \rightarrow \gamma + p.$$

   At recombination, the electrons and protons form neutral hydrogen atoms and the scattering partners disappear.

A third possibility is thermal freeze-out. It is similar to photon decoupling. Instead of $\gamma + e^- \rightarrow \gamma + e^-$ or $(\gamma + p \rightarrow \gamma + p)$ we now consider $X + \bar{X} \rightarrow Y + \bar{Y}$. While in the case of photon decoupling the basic reason was that the scattering partners fade away, the mechanism for thermal freeze-out is a little bit more subtle: In an expanding universe it becomes more and more unlikely for two particles X and $\bar{X}$ to find each other and to annihilate. This happens, when the annihilation rate

$$\Gamma_{X,\mathrm{eq}} = n_{X,\mathrm{eq}} \left\langle \sigma_{X\bar{X} \rightarrow Y\bar{Y}} \, v_{\mathrm{Møller}} \right\rangle,$$

where $n_{X,\mathrm{eq}}$ is the number density of particle X in equilibrium and $\left\langle \sigma_{X\bar{X} \rightarrow Y\bar{Y}} \, v_{\mathrm{Møller}} \right\rangle$ the thermal average of the annihilation cross section times velocity, becomes smaller than the Hubble parameter. Thus the condition for thermal freeze-out is

$$\Gamma_{X,\mathrm{eq}} = H.$$

Note that both sides have units s^{-1}.

10.2.1 The Boltzmann equation

Let us denote by $f_X(\vec{x}, \vec{p}, t)$ the phase space density of particle X.

$$f_X(\vec{x}, \vec{p}, t) \frac{d^3x \, d^3p}{(2\pi\hbar)^3}$$

gives the probability of finding a particle X at time t in a small volume $d^3x\, d^3p$ of phase space at the point $(\vec{x},\vec{p})$ in phase space.

The number density $n_X(\vec{x},t)$ is the integral of the phase space density over all momenta times a factor g_X^{spin}, taking degenerate states (e.g. spin states) into account:

$$n_X(\vec{x},t) = g_X^{\mathrm{spin}} \int \frac{d^3p}{(2\pi\hbar)^3} f_X(\vec{x},\vec{p},t).$$

The energy density is given by

$$\rho_X(\vec{x},t) = g_X^{\mathrm{spin}} \int \frac{d^3p}{(2\pi\hbar)^3} \sqrt{c^2\vec{p}^2 + c^4m^2}\, f_X(\vec{x},\vec{p},t).$$

The Boltzmann equation in classical statistical mechanics reads

$$\frac{d}{dt} f_X = \hat{C} f_X.$$

df_X/dt is called the flow term, $\hat{C}f_X$ is called the collision term. For the flow term we have

$$\frac{d}{dt} f_X = \frac{\partial}{\partial t} f_X + \frac{\partial \vec{x}}{\partial t}\vec{\nabla}_x f_X + \frac{\partial \vec{p}}{\partial t}\vec{\nabla}_p f_X.$$

Let us define an operator $\hat{L}$ by

$$\hat{L} = \frac{\partial}{\partial t} + \frac{\partial \vec{x}}{\partial t}\vec{\nabla}_x + \frac{\partial \vec{p}}{\partial t}\vec{\nabla}_p,$$

such that the left-hand side of the Boltzmann equation is $\hat{L}f_X$. We call $\hat{L}$ the **Liouville operator**.

Warning: In statistical mechanics the Liouville operator is usually defined slightly differently: Without the partial time derivative and for a N-particle system:

$$\hat{L}_{\text{statistical mechanic}} = \sum_{i=1}^{N} \frac{\partial \vec{x}_i}{\partial t}\vec{\nabla}_{x_i} + \frac{\partial \vec{p}_i}{\partial t}\vec{\nabla}_{p_i}.$$

Let us seek a generalisation of the Liouville operator to curved space. First of all, we write $f_X(\vec{x},\vec{p},t)$ as $f_X(x^\mu, p^\mu)$. Please note that both versions depend on seven independent variables, p^μ is constrained by

$$p_\mu p^\mu = m^2c^2.$$

Instead of the total time derivative (which would not respect general covariance) we consider the derivative with respect to an affine parameter λ:

$$\frac{d}{d\lambda} f_X = \frac{dx^\mu}{d\lambda}\frac{\partial f_X}{\partial x^\mu} + \frac{dp^\mu}{d\lambda}\frac{\partial f_X}{\partial p^\mu}.$$

From the geodesic equation we have

$$\frac{dp^\mu}{d\lambda} = -\Gamma^\mu_{\tau\sigma} p^\tau p^\sigma$$

and hence

$$\frac{d}{d\lambda} f_X = \left[p^\mu \frac{\partial}{\partial x^\mu} - \Gamma^\mu_{\tau\sigma} p^\tau p^\sigma \frac{\partial}{\partial p^\mu} \right] f_X.$$

Thus the generalisation of the Liouville operator is given by

$$\hat{L} = p^\mu \frac{\partial}{\partial x^\mu} - \Gamma^\mu_{\tau\sigma} p^\tau p^\sigma \frac{\partial}{\partial p^\mu}.$$

Let us now specialise to the Robertson-Walker metric. In a homogeneous and isotropic universe the phase space density $f_X(\vec{x},\vec{p},t)$ depends only on $E = \sqrt{c^2\vec{p}^2 + c^4 m^2}$ and t. Thus we consider $f_X(E,t)$. We obtain

$$\hat{L} f_X = \frac{E}{c^2} \frac{\partial}{\partial t} f_X - H \vec{p}^2 \frac{\partial}{\partial E} f_X,$$

or

$$\frac{c^2}{E} \hat{L} f_X = \frac{\partial}{\partial t} f_X - \frac{H c^2 \vec{p}^2}{E} \frac{\partial}{\partial E} f_X.$$

Our basic interest is the number density $n_X(t)$. We integrate the above equation over $\vec{p}$. We have

$$g_X^{\text{spin}} \int \frac{d^3 p}{(2\pi\hbar)^3} \frac{c^2}{E} \hat{L} f_X(E,t) = \frac{\partial}{\partial t} n_X - g_X^{\text{spin}} \int \frac{d^3 p}{(2\pi\hbar)^3} \frac{H c^2 \vec{p}^2}{E} \frac{\partial}{\partial E} f_X(E,t).$$

We simplify the second term with the help of integration-by-parts:

$$\begin{aligned}
-g_X^{\text{spin}} \int \frac{d^3 p}{(2\pi\hbar)^3} \frac{H c^2 \vec{p}^2}{E} \frac{\partial}{\partial E} f_X(E,t) &= -g_X^{\text{spin}} H c^2 \int \frac{dp d\Omega}{(2\pi\hbar)^3} \frac{p^4}{E} \frac{\partial}{\partial E} f_X(E,t) \\
&= -g_X^{\text{spin}} H \int \frac{dE d\Omega}{(2\pi\hbar)^3} \left(\frac{E^2}{c^2} - c^2 m^2 \right)^{\frac{3}{2}} \frac{\partial}{\partial E} f_X(E,t) \\
&= 3 g_X^{\text{spin}} H \int \frac{dE d\Omega}{(2\pi\hbar)^3} \frac{E p}{c^2} f_X(E,t) \\
&= 3 g_X^{\text{spin}} H \int \frac{dp d\Omega}{(2\pi\hbar)^3} p^2 f_X(E,t) \\
&= 3 g_X^{\text{spin}} H \int \frac{d^3 p}{(2\pi\hbar)^3} f_X(E,t) \; = \; 3 H n_X(t).
\end{aligned}$$

Thus

$$g_X^{\text{spin}} \int \frac{d^3p}{(2\pi\hbar)^3} \frac{c^2}{E} \hat{L} f_X(E,t) \;=\; \frac{\partial}{\partial t} n_X + 3Hn_X.$$

Let us now consider the collision term. We consider the processes $X + \bar{X} \to Y + \bar{Y}$ and $Y + \bar{Y} \to X + \bar{X}$. Integrated over the momenta the collision term is given by

$$\begin{aligned} & g_X^{\text{spin}} \int \frac{d^3p_X}{(2\pi\hbar)^3} \frac{c^2}{E_X} \hat{C} f_X(E_X,t) = \\ &= -c \sum_{\text{spins}} \int \frac{c\, d^3p_X}{(2\pi\hbar)^3 2E_X} \frac{c\, d^3p_{\bar{X}}}{(2\pi\hbar)^3 2E_{\bar{X}}} \frac{c\, d^3p_Y}{(2\pi\hbar)^3 2E_Y} \frac{c\, d^3p_{\bar{Y}}}{(2\pi\hbar)^3 2E_{\bar{Y}}} (2\pi\hbar)^4 \delta^4(p_X + p_{\bar{X}} - p_Y - p_{\bar{Y}}) \\ & \quad \times \left[f_X f_{\bar{X}} (1 \pm f_Y)(1 \pm f_{\bar{Y}}) |\mathcal{A}_{X\bar{X}\to Y\bar{Y}}|^2 - f_Y f_{\bar{Y}} (1 \pm f_X)(1 \pm f_{\bar{X}}) |\mathcal{A}_{Y\bar{Y}\to X\bar{X}}|^2 \right]. \end{aligned}$$

$\mathcal{A}_{X\bar{X}\to Y\bar{Y}}$ is the scattering amplitude for $X + \bar{X} \to Y + \bar{Y}$. If we normalise the creation and annihilation operators by

$$\left[\hat{a}_{\vec{p}}, \hat{a}^\dagger_{\vec{q}}\right] \;=\; (2\pi\hbar)^3 \delta^3(\vec{p} - \vec{q}),$$

the one-particle states by

$$|p\rangle \;=\; \sqrt{2\frac{E_{\vec{p}}}{c}} \hat{a}^\dagger_{\vec{p}} |0\rangle,$$

define the transition operator $\hat{T}$ by

$$\hat{S} \;=\; \mathbf{1} + i(2\pi\hbar)^4 \delta^4\left(\sum_{i=1}^n p_i\right) \hat{T},$$

and the n-particle scattering amplitude $\mathcal{A}_n(p_1, \dots, p_n)$ by

$$\left\langle 0 \left| i\hat{T} \right| p_1 \dots p_n \right\rangle \;=\; i\mathcal{A}_n(p_1, \dots, p_n)$$

and demand that the S-matrix operator $\hat{S}$ is dimensionless, we find that

$$\dim \mathcal{A}_n \;=\; [\dim p]^{4-n} [\dim \hbar]^{\frac{3}{2}n-4} \;=\; \left[\dim \frac{p}{\hbar}\right]^{4-n} [\dim \hbar]^{\frac{n}{2}}.$$

The factors $(1 \pm f_i)$ are of statistical origin and incorporate Bose enhancement $(1 + f_i)$ for bosons and Pauli blocking $(1 - f_i)$ for fermions.

We will make a few simplifying assumptions: We will assume that for all particles we have $E - \mu \gg k_B T$. In this limit the Bose-Einstein distribution and the Fermi-Dirac distribution reduce to the Maxwell-Boltzmann distribution:

$$\lim_{E-\mu \gg k_B T} \frac{1}{e^{\frac{1}{k_B T}(E-\mu)} \mp 1} \;=\; e^{-\frac{1}{k_B T}(E-\mu)}.$$

In this limit we may also neglect the statistical factors $(1 \pm f_i)$.

Secondly, we assume that the fundamental interactions entering the amplitude are T-invariant. This implies

$$\sum_{\text{spins}} |\mathcal{A}_{X\bar{X}\to Y\bar{Y}}|^2 = \sum_{\text{spins}} |\mathcal{A}_{Y\bar{Y}\to X\bar{X}}|^2 .$$

Thirdly, we assume that particles Y and $\bar{Y}$ go quickly into thermal equilibrium. This allows us to replace f_Y and $f_{\bar{Y}}$ by the equilibrium distributions

$$f_{Y,\text{eq}} = e^{-\frac{1}{k_B T}(E_Y-\mu_Y)}, \qquad f_{\bar{Y},\text{eq}} = e^{-\frac{1}{k_B T}(E_{\bar{Y}}-\mu_{\bar{Y}})}.$$

Fourthly, we neglect the chemical potentials. Due to the presence of the delta distribution $\delta(E_x + E_{\bar{X}} - E_Y - E_{\bar{Y}})$ we have

$$f_{X,\text{eq}} f_{\bar{X},\text{eq}} = e^{-\frac{1}{k_B T}(E_X+E_{\bar{X}})} = e^{-\frac{1}{k_B T}(E_Y+E_{\bar{Y}})} = f_{Y,\text{eq}} f_{\bar{Y},\text{eq}}.$$

Putting all this together, the collision term simplifies to

$$\begin{aligned}
& g_X^{\text{spin}} \int \frac{d^3 p_X}{(2\pi\hbar)^3} \frac{c^2}{E_X} \hat{C} f_X(E_X,t) = \\
& = -c \sum_{\text{spins}} \int \frac{c\, d^3 p_X}{(2\pi\hbar)^3\, 2E_X} \frac{c\, d^3 p_{\bar{X}}}{(2\pi\hbar)^3\, 2E_{\bar{X}}} \frac{c\, d^3 p_Y}{(2\pi\hbar)^3\, 2E_Y} \frac{c\, d^3 p_{\bar{Y}}}{(2\pi\hbar)^3\, 2E_{\bar{Y}}} (2\pi\hbar)^4 \delta^4\left(p_X + p_{\bar{X}} - p_Y - p_{\bar{Y}}\right) \\
& \quad \times |\mathcal{A}_{X\bar{X}\to Y\bar{Y}}|^2 \left(f_X f_{\bar{X}} - f_{X,\text{eq}} f_{\bar{X},\text{eq}} \right).
\end{aligned}$$

We introduce the cross section

$$\begin{aligned}
\sigma_{X\bar{X}\to Y\bar{Y}} = & \frac{1}{4\sqrt{(p_X \cdot p_{\bar{X}})^2 - c^4 m_X^2 m_{\bar{X}}^2}\, g_X^{\text{spin}} g_{\bar{X}}^{\text{spin}}} \sum_{\text{spins}} \int \frac{c\, d^3 p_Y}{(2\pi\hbar)^3\, 2E_Y} \frac{c\, d^3 p_{\bar{Y}}}{(2\pi\hbar)^3\, 2E_{\bar{Y}}} \\
& \times (2\pi\hbar)^4 \delta^4\left(p_X + p_{\bar{X}} - p_Y - p_{\bar{Y}}\right) |\mathcal{A}_{X\bar{X}\to Y\bar{Y}}|^2 .
\end{aligned}$$

Thus

$$\begin{aligned}
& g_X^{\text{spin}} \int \frac{d^3 p_X}{(2\pi\hbar)^3} \frac{c^2}{E_X} \hat{C} f_X(E_X,t) = \\
& \quad -g_X^{\text{spin}} g_{\bar{X}}^{\text{spin}} \int \frac{d^3 p_X}{(2\pi\hbar)^3} \frac{d^3 p_{\bar{X}}}{(2\pi\hbar)^3} \sigma_{X\bar{X}\to Y\bar{Y}} v_{\text{Møller}} \left(f_X f_{\bar{X}} - f_{X,\text{eq}} f_{\bar{X},\text{eq}} \right),
\end{aligned}$$

where $v_{\text{Møller}}$ is defined by

$$v_{\text{Møller}} = \frac{c^3 \sqrt{(p_X \cdot p_{\bar{X}})^2 - c^4 m_X^2 m_{\bar{X}}^2}}{E_X E_{\bar{X}}}.$$

Let us introduce the thermal average of the annihilation cross section times the velocity

$$\left\langle \sigma_{X\bar{X}\to Y\bar{Y}}\; v_{\text{Møller}} \right\rangle \;=\; \frac{g_X^{\text{spin}} g_{\bar{X}}^{\text{spin}}}{n_{X,\text{eq}} n_{\bar{X},\text{eq}}} \int \frac{d^3 p_X}{(2\pi\hbar)^3} \frac{d^3 p_{\bar{X}}}{(2\pi\hbar)^3} \sigma_{X\bar{X}\to Y\bar{Y}} v_{\text{Møller}} e^{-\frac{1}{k_B T}\left(E_X + E_{\bar{X}}\right)}$$

with

$$n_{X,\text{eq}} \;=\; g_X^{\text{spin}} \int \frac{d^3 p_X}{(2\pi\hbar)^3} e^{-\frac{E_X}{k_B T}},$$

and a similar definition applies to $n_{\bar{X},\text{eq}}$. We would like to express the collision term in terms of the thermal average of the annihilation cross section times the velocity. This is possible, if the phase space densities f_X and $f_{\bar{X}}$ are proportional to their thermal equilibrium densities $f_{X,\text{eq}}$ and $f_{\bar{X},\text{eq}}$ with a momentum-independent constant of proportionality. It can be shown that this is the case if the particles X and $\bar{X}$ stay after decoupling (when they are no longer in chemical equilibrium) in kinetic equilibrium. With this assumption one obtains for the collision term

$$g_X^{\text{spin}} \int \frac{d^3 p_X}{(2\pi\hbar)^3} \frac{c^2}{E_X} \hat{C} f_X(E_X, t) \;=\; -\left\langle \sigma_{X\bar{X}\to Y\bar{Y}}\; v_{\text{Møller}} \right\rangle \left(n_X n_{\bar{X}} - n_{X,\text{eq}} n_{\bar{X},\text{eq}} \right).$$

With $n_{\bar{X}} = n_X$ and $n_{\bar{X},\text{eq}} = n_{X,\text{eq}}$ we finally obtain the Boltzmann equation in a form most useful for cosmology:

$$\frac{\partial}{\partial t} n_X \;=\; -3 H n_X - \left\langle \sigma_{X\bar{X}\to Y\bar{Y}}\; v_{\text{Møller}} \right\rangle \left(n_X^2 - n_{X,\text{eq}}^2 \right).$$

We define the equilibrium annihilation rate as

$$\Gamma_{X,\text{eq}} \;=\; n_{X,\text{eq}} \left\langle \sigma_{X\bar{X}\to Y\bar{Y}}\; v_{\text{Møller}} \right\rangle.$$

10.2.2 The thermal average of the cross section times velocity

Let us work out in more detail n_{eq} and $\langle \sigma\, v_{\text{Møller}} \rangle$. We continue to work with Maxwell-Boltzmann distributions (and thus neglect differences between bosons and fermions). However, we allow for non-zero particle masses.

We start with n_{eq}. We have (g^{spin} denotes the degeneracy factor, i.e. the number of spin states) with $E = \sqrt{c^2 \vec{p}^2 + c^4 m^2}$

$$\begin{aligned}
n_{\text{eq}} &= g^{\text{spin}} \int \frac{d^3 p}{(2\pi\hbar)^3} e^{-\frac{E}{k_B T}} = \frac{4\pi g^{\text{spin}}}{(2\pi\hbar)^3} \int\limits_0^\infty dp\, p^2 e^{-\frac{E}{k_B T}} = \frac{4\pi g^{\text{spin}}}{c^3 (2\pi\hbar)^3} \int\limits_{mc^2}^\infty dE\, E \sqrt{E^2 - c^4 m^2}\, e^{-\frac{E}{k_B T}} \\
&= \frac{4\pi g^{\text{spin}}}{3 c^3 (2\pi\hbar)^3} \int\limits_{mc^2}^\infty dE \left[\frac{d}{dE} \left(E^2 - c^4 m^2 \right)^{\frac{3}{2}} \right] e^{-\frac{E}{k_B T}} \\
&= \frac{4\pi g^{\text{spin}}}{3 c^3 (2\pi\hbar)^3 k_B T} \int\limits_{mc^2}^\infty dE \left(E^2 - c^4 m^2 \right)^{\frac{3}{2}} e^{-\frac{E}{k_B T}}.
\end{aligned}$$

Let us now substitute

$$x = \frac{mc^2}{k_B T}, \qquad z = \frac{E}{mc^2}.$$

We obtain

$$n_{\text{eq}} = \frac{4\pi g^{\text{spin}} \left(mc^2\right)^4}{3c^3 \left(2\pi\hbar\right)^3 k_B T} \int\limits_1^\infty dz \left(z^2 - 1\right)^{\frac{3}{2}} e^{-xz}.$$

The modified Bessel function $K_\nu(x)$ is defined by

$$K_\nu(x) = \frac{\sqrt{\pi}}{\Gamma\left(\nu + \frac{1}{2}\right)} \left(\frac{x}{2}\right)^\nu \int\limits_1^\infty dz \left(z^2 - 1\right)^{\nu - \frac{1}{2}} e^{-xz},$$

and in particular

$$K_2(x) = \frac{x^2}{3} \int\limits_1^\infty dz \left(z^2 - 1\right)^{\frac{3}{2}} e^{-xz}.$$

Therefore

$$n_{\text{eq}} = \frac{4\pi g^{\text{spin}} \left(mc^2\right)^2 \left(k_B T\right)}{c^3 \left(2\pi\hbar\right)^3} K_2\left(\frac{mc^2}{k_B T}\right).$$

In the limit $mc^2 \ll k_B T$ (i.e. for relativistic particles) we have $x \to 0$. The modified Bessel function behaves as

$$K_2(x) \sim \frac{2}{x^2},$$

and we obtain

$$n_{\text{eq}} = \frac{8\pi g^{\text{spin}} \left(k_B T\right)^3}{c^3 \left(2\pi\hbar\right)^3} = g^{\text{spin}} \frac{\left(k_B T\right)^3}{\pi^2 \left(\hbar c\right)^3}$$

In the limit $mc^2 \gg k_B T$ (i.e. for non-relativistic particles) we have $x \to \infty$. The modified Bessel function behaves as

$$K_2(x) \sim \sqrt{\frac{\pi}{2x}} e^{-x}$$

and we obtain

$$n_{\text{eq}} = \frac{4\pi^{\frac{3}{2}} g^{\text{spin}} \left(mc^2\right)^{\frac{3}{2}} \left(k_B T\right)^{\frac{3}{2}}}{\sqrt{2} c^3 \left(2\pi\hbar\right)^3} e^{-\frac{mc^2}{k_B T}} = g^{\text{spin}} \left(\frac{m k_B T}{2\pi\hbar^2}\right)^{\frac{3}{2}} e^{-\frac{mc^2}{k_B T}}.$$

Let us now consider the thermal average of the cross section times velocity $\langle \sigma\, v_{\text{Møller}} \rangle$. We recall

$$\langle \sigma\, v_{\text{Møller}} \rangle \;=\; \frac{g_X^{\text{spin}} g_{\bar{X}}^{\text{spin}}}{n_{X,\text{eq}} n_{\bar{X},\text{eq}}} \int \frac{d^3 p_X}{(2\pi\hbar)^3} \frac{d^3 p_{\bar{X}}}{(2\pi\hbar)^3} \sigma v_{\text{Møller}} e^{-\frac{1}{k_B T}\left(E_X + E_{\bar{X}}\right)}.$$

The Mandelstam variable s is given by

$$s \;=\; \left(p_X + p_{\bar{X}}\right)^2 \;=\; p_X^2 + p_{\bar{X}}^2 + 2 p_X \cdot p_{\bar{X}} \;=\; c^2 m_X^2 + c^2 m_{\bar{X}}^2 + \frac{2}{c^2} E_X E_{\bar{X}} - 2\left|\vec{p}_X\right| \cdot \left|\vec{p}_{\bar{X}}\right| \cos\theta.$$

The cross section σ is a function of s. We examine the integral

$$\begin{aligned}
I &= \int \frac{d^3 p_X}{(2\pi\hbar)^3} \frac{d^3 p_{\bar{X}}}{(2\pi\hbar)^3} \sigma v_{\text{Møller}} e^{-\frac{1}{k_B T}\left(E_X + E_{\bar{X}}\right)} \\
&= \frac{8\pi^2}{(2\pi\hbar)^6} \int\limits_0^\infty dp_X \int\limits_0^\infty dp_{\bar{X}} \int\limits_0^\pi d\theta \sin\theta\, p_X^2 p_{\bar{X}}^2 \sigma v_{\text{Møller}} e^{-\frac{1}{k_B T}\left(E_X + E_{\bar{X}}\right)} \\
&= \frac{8\pi^2}{(2\pi\hbar)^6 c^4} \int\limits_{m_X c^2}^\infty dE_X \int\limits_{m_{\bar{X}} c^2}^\infty dE_{\bar{X}} \int\limits_0^\pi d\theta \sin\theta\, p_X E_X p_{\bar{X}} E_{\bar{X}} \sigma v_{\text{Møller}} e^{-\frac{1}{k_B T}\left(E_X + E_{\bar{X}}\right)}.
\end{aligned}$$

Let us now substitute the variable θ by the Mandelstam variable s. We have

$$\frac{ds}{d\theta} \;=\; 2 p_X p_{\bar{X}} \sin\theta.$$

We define

$$s_\pm \;=\; c^2 m_X^2 + c^2 m_{\bar{X}}^2 + \frac{2}{c^2} E_X E_{\bar{X}} \pm 2 p_X p_{\bar{X}}.$$

Thus

$$I \;=\; \frac{4\pi^2}{(2\pi\hbar)^6 c^4} \int\limits_{m_X c^2}^\infty dE_X \int\limits_{m_{\bar{X}} c^2}^\infty dE_{\bar{X}} \int\limits_{s_-}^{s_+} ds E_X E_{\bar{X}} \sigma v_{\text{Møller}} e^{-\frac{1}{k_B T}\left(E_X + E_{\bar{X}}\right)}.$$

We introduce

$$E_+ \;=\; E_X + E_{\bar{X}}, \qquad E_- \;=\; E_X - E_{\bar{X}}.$$

We change variables from $(E_X, E_{\bar{X}})$ to (E_+, E_-). In addition, we change the order of integration to (s, E_+, E_-). Let us work out the region of integration. From the Schwartz inequality we have

$$2 p_X \cdot p_{\bar{X}} \;\ge\; 2 m_X m_{\bar{X}} c^2$$

and therefore

$$s \;\ge\; \left(m_X + m_{\bar{X}}\right)^2 c^2.$$

The original constraints are

$$\begin{aligned} E_X &\geq m_X c^2, \\ E_{\bar{X}} &\geq m_{\bar{X}} c^2, \\ \left(s - c^2 m_X^2 - c^2 m_{\bar{X}}^2 - \frac{2}{c^2} E_X E_{\bar{X}}\right)^2 &\leq \frac{4}{c^4}\left(E_X^2 - c^4 m_X^2\right)\left(E_{\bar{X}}^2 - c^4 m_{\bar{X}}^2\right). \end{aligned}$$

The first two constraints give

$$E_+ \geq \left(m_X + m_{\bar{X}}\right) c^2, \qquad 2 m_X c^2 - E_+ \leq E_- \leq E_+ - 2 m_{\bar{X}} c^2.$$

We write the last constraint as a quadratic equation in E_-. For real solutions E_- the discriminant should be positive, this gives the constraint

$$E_+ \geq c\sqrt{s}.$$

Since $s \geq (m_X + m_{\bar{X}})^2 c^2$ the constraint $E_+ \geq (m_X + m_{\bar{X}})\, c^2$ is automatically satisfied. The solutions for E_- are

$$E_-^{\max/\min} = \frac{c}{s}\left(E_+\left(m_X^2 - m_{\bar{X}}^2\right) c \pm \sqrt{\left[s - c^2\left(m_X + m_{\bar{X}}\right)^2\right]\left[s - c^2\left(m_X - m_{\bar{X}}\right)^2\right]\left[\frac{E_+^2}{c^2} - s\right]}\right).$$

One checks that $E_-^{\max} \leq E_+ - 2 m_{\bar{X}} c^2$ and $E_-^{\min} \geq 2 m_X c^2 - E_+$. The requirement $E_-^{\max} \leq E_+ - 2 m_{\bar{X}} c^2$ is equivalent to

$$\left(2 m_{\bar{X}} E_+ - s + m_X^2 - m_{\bar{X}}^2\right)^2 \geq 0,$$

which for real values is always satisfied. The requirement $E_-^{\min} \geq 2 m_X c^2 - E_+$ leads to a similar condition, where m_X and m_{X} are exchanged. Thus

$$I = \frac{\pi^2}{(2\pi\hbar)^6 c} \int\limits_{(m_X + m_{\bar{X}})^2 c^2}^{\infty} ds \int\limits_{c\sqrt{s}}^{\infty} dE_+ \int\limits_{E_-^{\min}}^{E_-^{\max}} dE_- \sqrt{\left[s - c^2\left(m_X + m_{\bar{X}}\right)^2\right]\left[s - c^2\left(m_X - m_{\bar{X}}\right)^2\right]}\, \sigma e^{-\frac{E_+}{k_B T}}.$$

The integration over E_- is trivial. One obtains

$$\begin{aligned} I = {} & \frac{2\pi^2}{(2\pi\hbar)^6} \int\limits_{(m_X + m_{\bar{X}})^2 c^2}^{\infty} \frac{ds}{s}\left[s - c^2\left(m_X + m_{\bar{X}}\right)^2\right]\left[s - c^2\left(m_X - m_{\bar{X}}\right)^2\right] \sigma \\ & \times \int\limits_{c\sqrt{s}}^{\infty} dE_+ \left(\frac{E_+^2}{c^2} - s\right)^{\frac{1}{2}} e^{-\frac{E_+}{k_B T}}. \end{aligned}$$

The integration over E_+ yields a modified Bessel function K_1. We obtain

$$I = \frac{2\pi^2 k_B T}{(2\pi\hbar)^6} \int\limits_{(m_X+m_{\bar{X}})^2 c^2}^{\infty} \frac{ds}{\sqrt{s}} \left[s - c^2 (m_X + m_{\bar{X}})^2\right] \left[s - c^2 (m_X - m_{\bar{X}})^2\right] K_1\left(\frac{c\sqrt{s}}{k_B T}\right) \sigma,$$

and therefore

$$\langle \sigma\, v_{\text{Møller}} \rangle = \\ \frac{2\pi^2 g_X^{\text{spin}} g_{\bar{X}}^{\text{spin}} k_B T}{(2\pi\hbar)^6 n_{X,\text{eq}} n_{\bar{X},\text{eq}}} \int\limits_{(m_X+m_{\bar{X}})^2 c^2}^{\infty} \frac{ds}{\sqrt{s}} \left[s - c^2 (m_X + m_{\bar{X}})^2\right] \left[s - c^2 (m_X - m_{\bar{X}})^2\right] K_1\left(\frac{c\sqrt{s}}{k_B T}\right) \sigma.$$

In the case where X and $\bar{X}$ are particle and anti-particle, we have $m_X = m_{\bar{X}}$ and $g_X^{\text{spin}} = g_{\bar{X}}^{\text{spin}}$ and our formula simplifies to

$$\begin{aligned} \langle \sigma\, v_{\text{Møller}} \rangle &= \frac{2\pi^2 \left(g_X^{\text{spin}}\right)^2 k_B T}{(2\pi\hbar)^6 n_{X,\text{eq}}^2} \int\limits_{4m_X^2 c^2}^{\infty} ds \sqrt{s} \left[s - 4c^2 m_X^2\right] K_1\left(\frac{c\sqrt{s}}{k_B T}\right) \sigma \\ &= \frac{1}{8c^2 m_X^4 k_B T \left(K_2\left(\frac{m_X c^2}{k_B T}\right)\right)^2} \int\limits_{4m_X^2 c^2}^{\infty} ds \sqrt{s} \left[s - 4c^2 m_X^2\right] K_1\left(\frac{c\sqrt{s}}{k_B T}\right) \sigma, \end{aligned}$$

or in natural units $c = \hbar = k_B = 1$:

$$\langle \sigma\, v_{\text{Møller}} \rangle = \frac{1}{8 m_X^4 T \left(K_2\left(\frac{m_X}{T}\right)\right)^2} \int\limits_{4m_X^2}^{\infty} ds \sqrt{s} \left[s - 4m_X^2\right] K_1\left(\frac{\sqrt{s}}{T}\right) \sigma.$$

10.2.3 The effective number of relativistic degrees of freedom

In this paragraph we introduce two effective numbers of relativistic degrees of freedom, g_* and $g_{*,S}$. The first (g_*) one enters the relation between energy density and temperature, the second one ($g_{*,S}$) enters the relation between scale factor and temperature. We now distinguish between bosons and fermions, using Bose-Einstein and Fermi-Dirac distributions, respectively. However, we neglect particle masses. In the relativistic limit this is justified.

Let's start with g_*. Let's consider a relativistic boson with g_i^{spin} spin degrees of freedom at temperature T_i. A typical example is a photon, where $g_{\text{photon}}^{\text{spin}} = 2$. If the relativistic boson is decoupled, its spectral energy density corresponds to the temperature T_i, which does not need to be the temperature of the other particle species. The spectral energy density is given by

$$u_i(\omega, T_i) = \frac{g_i^{\text{spin}} \hbar}{2\pi^2 c^3} \frac{\omega^3}{e^{\frac{\hbar\omega}{k_B T_i}} - 1},$$

and the energy density is obtained by

$$\rho_i(T_i) = \int_0^\infty d\omega\, u_i(\omega, T_i) = \frac{g_i^{\text{spin}}}{2\pi^2}\frac{(k_B T_i)^4}{(\hbar c)^3}\int_0^\infty dx \frac{x^3}{e^x - 1} = g_i^{\text{spin}}\frac{\pi^2 (k_B T_i)^4}{30 (\hbar c)^3}.$$

This is the Stefan-Boltzmann law. For the number density we obtain

$$n_i(T_i) = \int_0^\infty d\omega\, \frac{u_i(\omega, T_i)}{\hbar\omega} = \frac{g_i^{\text{spin}}}{2\pi^2}\frac{(k_B T_i)^3}{(\hbar c)^3}\int_0^\infty dx \frac{x^2}{e^x - 1} = g_i^{\text{spin}}\frac{\zeta_3 (k_B T_i)^3}{\pi^2 (\hbar c)^3}.$$

Let's repeat the calculation for a relativistic fermion with g_i^{spin} spin degrees of freedom The spectral energy density is now

$$u_i(\omega, T_i) = \frac{g_i^{\text{spin}}\hbar}{2\pi^2 c^3}\frac{\omega^3}{e^{\frac{\hbar\omega}{k_B T_i}} + 1},$$

and we obtain for the energy density

$$\rho_i(T_i) = \int_0^\infty d\omega\, u_i(\omega, T_i) = \frac{g_i^{\text{spin}}}{2\pi^2}\frac{(k_B T_i)^4}{(\hbar c)^3}\int_0^\infty dx \frac{x^3}{e^x + 1} = \frac{7}{8} g_i^{\text{spin}}\frac{\pi^2 (k_B T_i)^4}{30 (\hbar c)^3}.$$

Compared to the boson case we get an extra factor $7/8$. For the number density we obtain

$$n_i(T_i) = \int_0^\infty d\omega\, \frac{u_i(\omega, T_i)}{\hbar\omega} = \frac{g_i^{\text{spin}}}{2\pi^2}\frac{(k_B T_i)^3}{(\hbar c)^3}\int_0^\infty dx \frac{x^2}{e^x + 1} = \frac{3}{4} g_i^{\text{spin}}\frac{\zeta_3 (k_B T_i)^3}{\pi^2 (\hbar c)^3}.$$

Compared to the boson case we get an extra factor $3/4$.

Let us now consider various relativistic species i, each with their own temperature T_i. The total energy density is then

$$\rho = \sum_i \rho_i(T_i).$$

Let us denote by T the photon temperature. We take T as a reference temperature. We may write

$$\rho = \frac{\pi^2}{30 (\hbar c)^3} g_* (k_B T)^4,$$

with

$$g_* = \sum_{\text{bosons}} g_i^{\text{spin}}\left(\frac{T_i}{T}\right)^4 + \frac{7}{8}\sum_{\text{fermions}} g_i^{\text{spin}}\left(\frac{T_i}{T}\right)^4.$$

This defines g_*. The effective number g_* enters the relation between the energy density and the temperature. The relation is applicable as long as the universe is radiation dominated (i.e. dominated by relativistic particles).

Let us now define $g_{*,S}$. The sub-script S refers to the entropy. We first consider an individual species of particles i. For vanishing chemical potential μ_i the entropy is given by

$$S_i \;=\; \frac{E_i + p_i V}{T_i},$$

where p_i denotes the pressure due to the species i. We will also consider the entropy density s_i:

$$s_i \;=\; \frac{S_i}{V} = \frac{\rho_i + p_i}{T_i}.$$

For relativistic particles we have

$$p_i \;=\; \frac{1}{3}\rho_i.$$

This holds for bosons and for fermions. To see this, we note that the pressure p_i is given for relativistic particles (with $E = cp$) by

$$\begin{aligned} p_i &= g_i^{\mathrm{spin}} \int \frac{d^3 p}{(2\pi\hbar)^3} \frac{c^2 \vec{p}^2}{3E} f_i(E, T_i) \;=\; 4\pi g_i^{\mathrm{spin}} \int\limits_0^\infty \frac{dp}{(2\pi\hbar)^3} \frac{c^2 p^4}{3E} f_i(E, T_i) \\ &= \frac{1}{3} \frac{g_i^{\mathrm{spin}} \hbar}{2\pi^2 c^3} \int\limits_0^\infty d\omega\, \omega^3 f_i(\hbar\omega, T_i) \;=\; \frac{1}{3} \int\limits_0^\infty d\omega\, u_i(\omega, T_i) \;=\; \frac{1}{3}\rho_i. \end{aligned}$$

Please note that p_i denotes the pressure, while $p = |\vec{p}|$ denotes the absolute value of the three-momentum. We further used

$$f_i(E, T_i) \;=\; \frac{1}{e^{\frac{E}{k_B T_i}} \mp 1}, \qquad u_i(\omega, T_i) \;=\; \frac{g_i^{\mathrm{spin}} \hbar}{2\pi^2 c^3} \frac{\omega^3}{e^{\frac{\hbar\omega}{k_B T_i}} \mp 1}.$$

The entropy density is therefore given by

$$s_i \;=\; \frac{4}{3} \frac{\rho_i}{T_i}.$$

Adding up the different species we obtain

$$s \;=\; \frac{2\pi^2 k_B}{45 (\hbar c)^3} g_{*,S} (k_B T)^3,$$

with

$$g_{*,S} \;=\; \sum_{\text{bosons}} g_i^{\mathrm{spin}} \left(\frac{T_i}{T}\right)^3 + \frac{7}{8} \sum_{\text{fermions}} g_i^{\mathrm{spin}} \left(\frac{T_i}{T}\right)^3.$$

If entropy is conserved, we have

$$g_{*,S}T^3R^3 \quad = \quad \text{const},$$

leading to

$$T \quad \sim \quad g_{*,S}^{-\frac{1}{3}}\frac{1}{R}.$$

10.3 Neutrinos and hot relics

Let us now discuss the implications of the Boltzmann equation

$$\frac{d}{dt}n_X \quad = \quad -3Hn_X - \langle \sigma\, v_{\text{Møller}} \rangle \left(n_X^2 - n_{X,\text{eq}}^2 \right).$$

The first term on the right-hand side gives the dilution of the number density due to the expansion of the universe. The second term accounts for annihilation, while the third term (which comes with a positive sign) corresponds to the production process. As long as the first term on the right-hand side can be neglected against the second and the third term, the Boltzmann equation will drive the number density n_X towards the equilibrium number density $n_{X,\text{eq}}$. This changes when the first term becomes comparable to the other two terms. We define the freeze-out condition by

$$n_{\text{eq}} \langle \sigma\, v_{\text{Møller}} \rangle \quad = \quad H.$$

For the thermal average of the cross section times velocity we will use very crude approximations. For relativistic particles we will assume

$$\langle \sigma\, v_{\text{Møller}} \rangle \quad = \quad \hbar^2 c^3 g^4 \frac{(k_B T)^2}{(m_{\text{mediator}} c^2)^4},$$

where g is a dimensionless coupling and m_{mediator} is the mass of a mediator particle through which annihilation proceeds.

An example for relativistic particles are **neutrinos**. In this case $m_{\text{mediator}} = m_Z$ and g is the weak coupling. Fermi's constant is defined by

$$\frac{G_F}{(\hbar c)^3} \quad = \quad \frac{\sqrt{2} e^2}{8 \sin^2 \theta_W m_W^2 c^4} \;\approx\; 1.166 \cdot 10^{-5} \text{GeV}^{-2}.$$

For the neutrino annihilation cross section we make the crude approximation

$$\langle \sigma\, v_{\text{Møller}} \rangle \quad = \quad \hbar^2 c^3 \frac{G_F^2}{(\hbar c)^6} (k_B T)^2.$$

For relativistic fermions we use

$$n_{\text{eq}} \quad = \quad \frac{3}{4} \frac{\zeta_3 (k_B T)^3}{\pi^2 (\hbar c)^3}.$$

We assume that decoupling of the neutrinos occurs, when the universe is radiation dominated. With the first Friedmann equation and the effective number g_* of relativistic degrees of freedom

$$H^2 = \frac{8\pi G}{3c^2}\rho, \qquad \rho = \frac{\pi^2}{30\left(\hbar c\right)^3} g_* \left(k_B T\right)^4$$

one obtains

$$H = \sqrt{\frac{4\pi^3 G g_*}{45\hbar^3 c^5}} \left(k_B T\right)^2.$$

We may now calculate the freeze-out temperature:

$$\begin{aligned} n_{\mathrm{eq}} \left\langle \sigma \, v_{\text{Møller}} \right\rangle &= H, \\ \frac{3\zeta_3 \left(k_B T\right)^3}{4\pi^2 \left(\hbar c\right)^3} \cdot \hbar^2 c^3 \frac{G_F^2}{\left(\hbar c\right)^6} \left(k_B T\right)^2 &= \sqrt{\frac{4\pi^3 G g_*}{45\hbar^3 c^5}} \left(k_B T\right)^2, \\ k_B T &= \pi \left(\frac{4}{3\zeta_3}\right)^{\frac{1}{3}} \left(\frac{G_F}{\left(\hbar c\right)^3}\right)^{-\frac{2}{3}} \left(\frac{4\pi G g_*}{45\hbar c^5}\right)^{\frac{1}{6}}. \end{aligned}$$

Let us first calculate the effective number g_* of relativistic degrees of freedom. Let us assume that the relativistic particles at freeze-out are photons, electrons, positrons and neutrinos. We have

	fermion factor	g^{spin}
γ		2
e^-	$\frac{7}{8}$	2
e^+	$\frac{7}{8}$	2
ν_e, ν_μ, ν_τ	$\frac{7}{8}$	1
$\bar{\nu}_e, \bar{\nu}_\mu, \bar{\nu}_\tau$	$\frac{7}{8}$	1
g_*		$\frac{43}{4}$

With $G = 6.7086 \cdot 10^{-39} \hbar c^5 \text{GeV}^{-2}$ one obtains an estimate for the freeze-out temperature of neutrinos:

$$k_B T \approx \pi \left(\frac{4}{3\zeta_3}\right)^{\frac{1}{3}} \left(1.166 \cdot 10^{-5}\right)^{-\frac{2}{3}} \left(\frac{4\pi}{45} \cdot \frac{43}{4} \cdot 6.7086 \cdot 10^{-39}\right)^{\frac{1}{6}} \text{GeV} \approx 3.3 \text{ MeV}.$$

With the current upper limit on the neutrino masses $m_\nu c^2 < 2$ eV we have

$$m_\nu c^2 \ll k_B T,$$

which justifies a posteriori the use of the relativistic approximation. It also justifies a posteriori that the relativistic degrees of freedom are photons, electrons, positrons and neutrinos. Neutrinos are hot relics.

Let us discuss the **temperature of the cosmic neutrino background**. Our previous formulae are valid as long as the neutrinos are relativistic. We denote by $T_{1,\nu}$ the temperature of the neutrinos at decoupling. At decoupling the neutrino temperature equals the temperature of the rest of the universe, and in particular equals the temperature of the photons $T_{1,\gamma}$:

$$T_{1,\nu} = T_{1,\gamma}.$$

After decoupling, the temperature of the neutrinos is simply red-shifted:

$$T_{2,\nu} = \left(\frac{R_1}{R_2}\right) T_{1,\nu},$$

where R_1 is the scale factor at decoupling and R_2 is the scale factor at time t_2. We are interested in the relation of the neutrino temperature to the photon temperature. We have seen that the neutrinos decouple around $k_B T \approx 3.3$ MeV. Around 1 MeV ($\approx 2 \cdot 511$ keV) a large fraction of electrons and positrons annihilate, leaving only a tiny fraction of electrons behind (which are part of the observed matter today). The electron-positron annihilation reheats the photon gas. We may calculate the change in the photon temperature due to reheating, assuming that the process conserves entropy. Let us introduce the effective number $g_{*,S}^{\mathrm{before}}$ corresponding to just before electron-positron annihilation and taking only photons, electrons and positrons into account. $g_{*,S}^{\mathrm{before}}$ is given by

$$g_{*,S}^{\mathrm{before}} = 2 + \frac{7}{8} \cdot 2 \cdot 2 = \frac{11}{2}.$$

Immediately after electron-positron annihilation there are only photons (and neutrinos) left and we set

$$g_{*,S}^{\mathrm{after}} = 2.$$

The neutrinos are already decoupled and take no part in the temperature/entropy increase. Their entropy is the same before and after electron-positron annihilation. We further assume that electron-positron annihilation occurs in a time interval, where we may neglect changes in the scale factor R. If the entropy is conserved (i.e. the entropy from the electrons/positrons is transferred to the photons) we have

$$g_{*,S}^{\mathrm{before}} \left(T^{\mathrm{before}} R\right)^3 + S_\nu = g_{*,S}^{\mathrm{after}} \left(T^{\mathrm{after}} R\right)^3 + S_\nu,$$

where S_ν denotes the entropy of the neutrinos. We therefore have

$$T^{\mathrm{after}} = \left(\frac{g_{*,S}^{\mathrm{before}}}{g_{*,S}^{\mathrm{after}}}\right)^{\frac{1}{3}} T^{\mathrm{before}}.$$

Thus after electron-positron annihilation the neutrino temperature and the photon temperature are related by

$$T_\nu = \left(\frac{g_{*,S}^{\mathrm{after}}}{g_{*,S}^{\mathrm{before}}}\right)^{\frac{1}{3}} T_\gamma = \left(\frac{4}{11}\right)^{\frac{1}{3}} T_\gamma.$$

As long as the neutrinos are relativistic (and after electron-positron annihilation) we have for the effective number $g_{*,S}$

$$g_{*,S} = 2+\frac{7}{8}\cdot 2\cdot 3\cdot \frac{4}{11} = \frac{43}{11} \approx 3.91.$$

This is based on the assumption that the neutrinos freeze-out first and electron-positron annihilation occurs afterwards. As the temperatures of neutrino freeze-out and electron-positron annihilation are quite close, the neutrino freeze-out is not fully completed as electron-positron annihilation starts. Thus some energy/entropy is transferred to the neutrinos. This leads to a small corrections, which may be described by changing the number of neutrinos from three to an effective number of neutrino species $N_{\mathrm{eff}} = 3.046$. This yields

$$g_* = 3.38, \quad g_{*,S} = 3.94.$$

Let us now discuss the **neutrino contribution to the density parameter**. We set

$$\Omega_\nu = \frac{8\pi G}{3c^2H^2}\rho_\nu,$$

Let us assume that the neutrinos have (small) masses and that they are non-relativistic today. Let t_2 denote today's time. With three neutrinos (and three anti-neutrinos) we have

$$\rho_\nu = 2\left(\sum_i m_i c^2\right) n_\nu(t_2),$$

where $n_\nu(t_2)$ denotes today's number density of one neutrino species. We assume the number densities of all neutrinos (and anti-neutrinos) to be the same. Let $n_\gamma(t_2)$ denote today's number density of the photons from the cosmic microwave background. $n_\gamma(t_2)$ is given by

$$n_\gamma(t_2) = \frac{2\zeta_3}{\pi^2}\frac{\left(k_B T_{\gamma,2}\right)^3}{(\hbar c)^3}.$$

We may re-write the energy density as

$$\rho_\nu = 2\left(\sum_i m_i c^2\right)\left(\frac{n_\nu(t_2)}{n_\gamma(t_2)}\right) n_\gamma(t_2).$$

Let us now consider a time t_1, where the neutrinos where still relativistic, but after neutrinos and photons decoupled. Since both neutrinos and photons are decoupled their numbers does not change from t_1 to t_2 and we have

$$\frac{n_\nu(t_1)}{n_\gamma(t_1)} = \frac{n_\nu(t_2)}{n_\gamma(t_2)}.$$

Since the neutrinos are still relativistic at t_1, we may use

$$n_\nu(t_1) = \frac{3\zeta_3}{4\pi^2}\frac{(k_B T_{\nu,1})^3}{(\hbar c)^3}.$$

Combining everything we obtain

$$\rho_\nu = \frac{3}{4}\left(\sum_i m_i c^2\right)\left(\frac{T_{\nu,1}}{T_{\gamma,1}}\right)^3 n_\gamma(t_2) = \frac{3}{11}\left(\sum_i m_i c^2\right) n_\gamma(t_2)$$

and thus

$$\Omega_\nu = \frac{8\pi G}{3c^2H^2}\cdot\frac{3}{11}\left(\sum_i m_i c^2\right) n_\gamma(t_2) = \frac{48\zeta_3}{33\pi}\left(\sum_i m_i c^2\right)\left(\frac{G}{c^5\hbar}\right)\frac{(k_B T_{\gamma,2})^3}{(\hbar H)^2}.$$

with $k_B T_{\gamma,2} = 2.35\cdot 10^{-4}$ eV and $\hbar H = 1.45\cdot 10^{-33}$ eV one obtains

$$\Omega_\nu \approx 0.023\left(\frac{\sum_i m_i c^2}{\text{eV}}\right).$$

The neutrinos are hot dark matter. They cannot constitute the bulk of dark matter. Let us assume that their contribution to the density parameter is x, with

$$x < \Omega_{DM} \approx 0.26.$$

We thus obtain a bound on the neutrino masses

$$\sum_i m_i c^2 < x\cdot 43.2\,\text{eV}.$$

A conservative estimate for x is $x < 0.13$, e.g. assuming that the neutrino contribution is not more than half of the total dark matter contribution. We then find

$$\sum_i m_i c^2 < 5.6\,\text{eV}.$$

10.4 Cold relics and the WIMP miracle

Let us now turn to **cold dark matter**. We consider a dark matter particle with mass m_X and we will assume that this particle decouples when it is non-relativistic, i.e. $m_X c^2 \gg k_B T$. We do not distinguish between bosons and fermions and work for simplicity with the Maxwell-Boltzmann distribution.

The thermal average of the annihilation cross section times velocity is denoted by $\langle\sigma\, v_{\text{Møller}}\rangle$. We keep the dependence on this quantity explicit. A concrete crude approximation is for example given by

$$\langle\sigma\, v_{\text{Møller}}\rangle = \beta\,\hbar^2 c^3\frac{G_F^2}{(\hbar c)^6}\left(m_X c^2\right)^2.$$

Such a cross section would arise, if the dark matter particles annihilate through a mediator particle with coupling and mass similar to the electro-weak bosons. This is encoded in the factor G_F^2. In order to get the dimensions right, we need an additional factor E^2. The appropriate scale is now $m_X c^2$. $\beta = v/c$ is the ratio of a typical non-relativistic velocity to the speed of light. As a numerical example let us take $m_X = 1$ TeV and

$$\langle \sigma \, v_{\text{Møller}} \rangle_{\text{ref}} \;\; = \;\; 1.4 \cdot 10^{-5} \, \cdot \, \hbar^2 c^3 \frac{G_F^2}{(\hbar c)^6} \left(m_X c^2 \right)^2 \, \approx \, 2.22 \cdot 10^{-32} \, \text{m}^3 \text{s}^{-1}.$$

For the number density we use the non-relativistic approximation

$$n_{\text{eq}} \;\; = \;\; \left(\frac{m_X k_B T}{2\pi\hbar^2} \right)^{\frac{3}{2}} e^{-\frac{m_X c^2}{k_B T}}.$$

We will again assume that freeze-out occurs while the universe is radiation dominated. Thus

$$H \;\; = \;\; \sqrt{\frac{4\pi^3 G g_*}{45 \hbar^3 c^5}} \left(k_B T \right)^2.$$

Assuming that the freeze-out occurs before electroweak symmetry breaking, we may assume that all particles of the Standard Model are relativistic. Thus

$$g_* \;\; = \;\; \underbrace{2\left(1+3+8\right)}_{\text{gauge bosons}} + \frac{7}{8} \left[\underbrace{2 \cdot 2 \cdot 6 \cdot 3}_{\text{quarks}} + \underbrace{2 \cdot 2 \cdot 3}_{\text{charged leptons}} + \underbrace{2 \cdot 3}_{\text{neutrinos}} \right] + \underbrace{4}_{\text{Higgs}} \;\; = \;\; \frac{427}{4}.$$

Before electroweak symmetry breaking the complex Higgs doublet contributes four degrees of freedom. Three degrees of freedom become after electroweak symmetry breaking the longitudinal modes of the $W^{\pm}$- and Z-bosons. The fourth degree of freedom is the Higgs boson. Let us introduce

$$x \;\; = \;\; \frac{m_X c^2}{k_B T}.$$

The condition for freeze-out

$$n_{\text{eq}} \left\langle \sigma \, v_{\text{Møller}} \right\rangle \;\; = \;\; H$$

yields

$$\begin{aligned}
\left(\frac{m_X c^2 k_B T}{2\pi \left(\hbar c\right)^2} \right)^{\frac{3}{2}} e^{-x} \left\langle \sigma \, v_{\text{Møller}} \right\rangle \;\; &= \;\; \sqrt{\frac{4\pi^3 G g_*}{45 \hbar^3 c^5}} \left(k_B T \right)^2, \\
\sqrt{x} e^{-x} \;\; &= \;\; \pi^3 \sqrt{\frac{32 G g_*}{45 \hbar c^5}} \frac{\hbar^2 c^3}{m_X c^2 \left\langle \sigma \, v_{\text{Møller}} \right\rangle}.
\end{aligned}$$

Given m_X and $\langle \sigma\, v_{\text{Møller}} \rangle$ we may solve this equation (numerically) for x and obtain in this way the freeze-out temperature T_1. For our numerical example $m_X = 1$ TeV and $\langle \sigma\, v_{\text{Møller}} \rangle_{\text{ref}} = 2.22 \cdot 10^{-32}$ m^{3}s^{-1} we find $x = 26.82$.

Let us now turn to the contribution to the density parameter. We denote quantities at the time of the freeze-out with a subscript 1, while today's quantities are denoted with a subscript 2. We first consider a crude "sudden approximation": For $T > T_1$ we assume that the number density n_X is given by the equilibrium distribution, while for $T < T_1$ we assume that all particles X are frozen out and the total number of particles X stays constant. With these assumptions we have

$$\Omega_X = \frac{8\pi G}{3c^2H^2}\rho_X(t_2), \qquad \rho_X(t_2) = m_X c^2 n_X(t_2), \qquad n_X(t_2) = \left(\frac{R_1}{R_2}\right)^3 n_X(t_1).$$

The last equation states that after freeze-out the dark matter particles X are decoupled. The number of particles X is conserved, the number density is diluted by the third power of the scale factor. For this factor we have

$$\left(\frac{R_1}{R_2}\right)^3 = \frac{g_{*,S,2}}{g_{*,S,1}}\left(\frac{T_2}{T_1}\right)^3,$$

with

$$g_{*,S,1} = \frac{427}{4}, \qquad g_{*,S,2} = 3.94$$

and $T_2 = 2.73K$ is the temperature of the cosmic microwave background. Putting everything together we get

$$\begin{aligned}
\Omega_X &= \frac{8\pi G}{3c^2H^2} m_X c^2 \frac{g_{*,S,2}}{g_{*,S,1}}\left(\frac{T_2}{T_1}\right)^3 \left(\frac{m_X k_B T_1}{2\pi\hbar^2}\right)^{\frac{3}{2}} e^{-\frac{m_X c^2}{k_B T_1}} \\
&= \frac{4}{3\sqrt{2\pi}(\hbar H)^2}\left(\frac{G}{\hbar c^5}\right)\left(m_X c^2\right)^4 \frac{g_{*,S,2}}{g_{*,S,1}}\left(\frac{T_2}{T_1}\right)^3 x_1^{-\frac{3}{2}} e^{-x_1} \\
&= \frac{\pi}{9\sqrt{10}}\sqrt{g_{*,1}}\frac{g_{*,S,2}}{g_{*,S,1}}\left(\frac{8\pi G}{\hbar c^5}\right)^{\frac{3}{2}} \frac{c^3 (k_B T_2)^3}{H^2 \langle \sigma\, v_{\text{Møller}} \rangle} x_1.
\end{aligned}$$

For our numerical example $m_X = 1$ TeV and $\langle \sigma\, v_{\text{Møller}} \rangle_{\text{ref}} = 2.22 \cdot 10^{-32}$ m^{3}s^{-1} (and $g_{*,1} = g_{*,S,1}$) we obtain

$$\Omega_X \approx 0.25.$$

In the discussion above we made the (unrealistic) assumption that above the freeze-out temperature the number density n_X is in thermal equilibrium, while below the freeze-out temperature the particle number N_X is constant. In reality freeze-out does nor occur suddenly, but proceeds gradually. We may model this more accurately with the help of the Boltzmann equation

$$\frac{\partial}{\partial t} n_X = -3Hn_X - \langle \sigma\, v_{\text{Møller}} \rangle \left(n_X^2 - n_{X,\text{eq}}^2\right).$$

It is convenient to use as evolution variable not the time t, but the dimensionless parameter $x = m_X c^2/(k_B T)$ introduced above. In addition, we scale out the effect of the expansion of the universe by considering instead of n_X the quantity

$$Y = \frac{n_X}{s},$$

where s is the entropy density. Note that $sR^3 = \text{const}$ and hence

$$\begin{aligned} R^3 \frac{\partial s}{\partial t} + 3sR^2\dot{R} &= 0, \\ \frac{\partial s}{\partial t} &= -3sH. \end{aligned}$$

Furthermore

$$\frac{\partial x}{\partial t} = -\frac{m_X c^2}{k_B T}\frac{1}{T}\frac{\partial T}{\partial t} = -\frac{m_X c^2}{k_B T}\frac{1}{3T^3}\frac{\partial T^3}{\partial t} = -\frac{m_X c^2}{k_B T}\frac{1}{3s}\frac{\partial s}{\partial t} = \frac{m_X c^2}{k_B T} H = xH.$$

The Boltzmann equation may re-written as

$$\frac{\partial}{\partial x} Y = -\frac{s}{xH}\left\langle \sigma\, v_{\text{Møller}} \right\rangle \left(Y^2 - Y_{\text{eq}}^2\right).$$

The Hubble parameter is x-dependent. In a radiation dominated universe we have $H^2 \sim \rho \sim T^4$. Thus

$$H = \frac{H(x=1)}{x^2}.$$

Let us introduce

$$\lambda = \frac{sx^3}{H(x=1)}\left\langle \sigma\, v_{\text{Møller}} \right\rangle = \frac{2\pi^2 k_B}{45(\hbar c)^3} g_{*,S} \frac{\left(m_X c^2\right)^3 \left\langle \sigma\, v_{\text{Møller}} \right\rangle}{H(x=1)}.$$

Assuming that in the range of interest $\langle \sigma\, v_{\text{Møller}} \rangle$ and $g_{*,S}$ are temperature-independent, it follows that λ is temperature-independent constant as well, and hence a x-independent constant. Then

$$\frac{\partial}{\partial x} Y = -\frac{\lambda}{x^2}\left(Y^2 - Y_{\text{eq}}^2\right).$$

At high temperatures, corresponding to $x \ll 1$ we have $Y \approx Y_{\text{eq}}$. This gives a boundary condition and we may integrate the differential equation numerically towards low temperatures, corresponding to $x \gg 1$.

In order to get a qualitative understanding we consider the following approximation: For $x \gg x_1$ (where x_1 denotes the freeze-out value defined by $\Gamma_{X,\text{eq}} = H$) we have $Y \gg Y_{\text{eq}}$ and the differential equation simplifies to

$$\frac{\partial}{\partial x} Y = -\frac{\lambda}{x^2} Y^2.$$

Integration from $x = x_1$ to $x = \infty$ yields

$$\frac{1}{Y_\infty} - \frac{1}{Y_1} = \frac{\lambda}{x_1}.$$

Typically, $Y_1 \gg Y_\infty$ and hence

$$Y_\infty = \frac{x_1}{\lambda}.$$

Within this approximation we obtain

$$\Omega_X = \frac{\pi}{9\sqrt{10}} \sqrt{g_{*,1}} \frac{g_{*,S,2}}{g_{*,S,1}} \left(\frac{8\pi G}{\hbar c^5}\right)^{\frac{3}{2}} \frac{c^3 \left(k_B T_2\right)^3}{H^2 \left\langle \sigma \, v_{\text{Møller}} \right\rangle} x_1,$$

i.e. the same result as within the “sudden approximation”. This is not surprising, as we made again essentially the same approximation. Please note the factor $1/\langle \sigma \, v_{\text{Møller}} \rangle$, the higher the thermal average of the cross section times velocity, the lower the relic abundance. The factor x_1 depends only mildly on the product $m_X \langle \sigma \, v_{\text{Møller}} \rangle$.

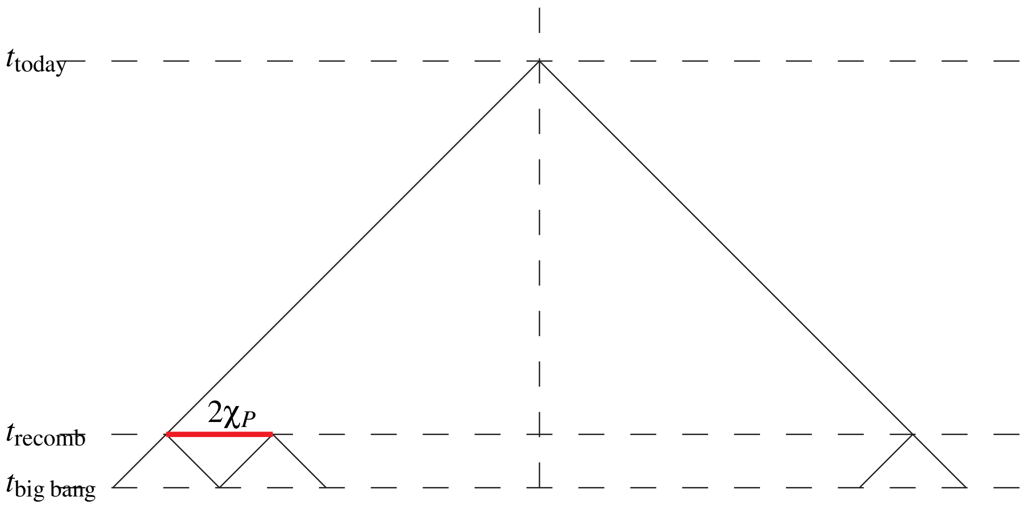
11 Inflation

11.1 The horizon problem

Let us consider photons from the cosmic microwave background. They decoupled at $t_1 = t_{\text{recomb}}$ and had no interaction afterwards. At t_1 their particle horizon is given by

$$\chi_p = c\int_{t_0}^{t_1} \frac{dt}{R(t)} = c\int_{R_0}^{R_1} \frac{dR}{R^2 H(R)} = \frac{c}{R_2 H_2}\int_{z_1}^{z_0} \frac{dz}{E(z)} = \frac{c}{R_2 H}\int_{z_1}^{\infty} \frac{dz}{E(z)},$$

where today's quantities ($t_2 = t_{\text{today}}$) are denoted by $H_2 = H$. Quantities at the time of the big bang ($t_0 = t_{\text{big bang}}$) are denoted with a subscript 0.



Up to recombination the universe was dominated by radiation and matter. We therefore model $E(z)$ by

$$E(z) = \left[\Omega_{R,2}(1+z)^4 + \Omega_{M,2}(1+z)^3\right]^{\frac{1}{2}}.$$

We obtain

$$\chi_p = \frac{2c}{R_2 H \Omega_{M,2}^{\frac{1}{2}} (1+z_1)^{\frac{1}{2}}}\left[\sqrt{1+(1+z_1)\frac{\Omega_{R,2}}{\Omega_{M,2}}} - \sqrt{(1+z_1)\frac{\Omega_{R,2}}{\Omega_{M,2}}}\right].$$

Numerically we have with $z_1 = 1100$, $\Omega_{R,2} = 10^{-5}$ and $\Omega_{M,2} = 0.31$

$$C_1 = \sqrt{1+(1+z_1)\frac{\Omega_{R,2}}{\Omega_{M,2}}} - \sqrt{(1+z_1)\frac{\Omega_{R,2}}{\Omega_{M,2}}} = 0.83.$$

Thus

$$2d_P = 2R_1\chi_P = \frac{4cR_1}{R_2 H \Omega_{M,2}^{\frac{1}{2}} (1+z_1)^{\frac{1}{2}}} C_1 = \frac{4c}{H \Omega_{M,2}^{\frac{1}{2}} (1+z_1)^{\frac{3}{2}}} C_1.$$

An event at $t_0 = t_{\text{big bang}}$ cannot influence simultaneously two photons, which were separated more than $2d_P$ at recombination time.

Let us now consider an object of spatial size d_{trans}, which is observed on the sky today ($t_2 = t_{\text{today}}$) and extends over an angle θ. The **angular diameter distance** d_A is defined by

$$d_A = \frac{d_{\text{trans}}}{\theta}.$$

It can be shown that the angular diameter distance is related to the luminosity distance

$$d_L = (1+z)^2 d_A,$$

where z is the red shift. We have

$$d_A = \frac{d_L}{(1+z_1)^2} = \frac{c}{H(1+z_1)} \int\limits_0^{z_1} \frac{dz}{E(z)}.$$

With

$$E(z) = \left[0.69 + 0.31(1+z)^3\right]^{\frac{1}{2}}$$

we obtain for the integral

$$\int\limits_0^{z_1} \frac{dz}{E(z)} \approx 3.15.$$

Let us now consider the angle under which we observe today a region of cosmic microwave photons, which could have had a chance to reach thermal equilibrium between t_0 and t_1. We have

$$\theta = \frac{2d_P}{d_A} \approx \frac{4}{3.15} \frac{1}{\sqrt{\Omega_M}} \frac{C_1}{\sqrt{1+z_1}} \approx 5.7 \cdot 10^{-2} \approx 3^\circ.$$

Further more

$$\frac{\theta^2}{4\pi} = 2.6 \cdot 10^{-4}.$$

We observe in experiments that the cosmic microwave background is isotropic over the complete sky with anisotropies $\leq 10^{-5}$. Within the Robertson-Walker model of cosmology we see that photons could have reached thermal equilibrium between $t_0 = t_{\text{big bang}}$ and $t_1 = t_{\text{recomb}}$ in regions of the size θ^2. Within these regions we would expect the cosmic microwave background to be isotropic. The fact that the cosmic microwave background is isotropic over the complete sky can be explained within the Robertson-Walker model of cosmology only by fine-tuned initial conditions at $t_0 = t_{\text{big bang}}$. This is the horizon problem.

11.2 The flatness problem

Consider the differential equation

$$\frac{dx}{dt} = \lambda (x - x_0)$$

with two constants λ and x_0. It is clear that $x(t) = x_0$ is a fixed-point of the differential equation. What happens if we start from the initial condition at time t_1 with initial value

$$x(t_1) = x_0 + \delta,$$

with δ small? For $\lambda < 0$ the system will evolve towards the fixed point and we call the point $x(t) = x_0$ a **stable fixed point**.

For $\lambda > 0$ the system will evolve away from the fixed point. We say that in this case the point $x(t) = x_0$ is an **unstable fixed point**.

Let us now assume $\lambda > 0$. A solution to the differential equation with initial condition $x(t_1) = x_0 + \delta$ is given by

$$x(t) = x_0 + \delta e^{\lambda(t-t_1)}.$$

Assume now that we observe today (at time t_1) the value $x_0 + \delta$. We may then ask, what was the initial condition at an earlier time t_0 leading to the observed value $x(t_1) = x_0 + \delta$ today. This is easily answered:

$$x(t_0) = x_0 + \delta e^{\lambda(t_0-t_1)}.$$

In other words, if the value today is a small quantity δ away from the unstable fixed point x_0, it must have even closer (by an exponential factor) to the unstable fixed point at earlier times:

$$x(t_0) - x_0 = \delta e^{-\lambda(t_1-t_0)}.$$

Thus we need very precisely fine-tuned initial conditions at time t_0 to explain the observed value at t_1 today.

Let us now apply this to cosmology. The time evolution of the density parameter

$$\Omega = 1 + \kappa \frac{c^2}{H^2 R^2}.$$

is given by

$$\frac{d}{dt}\Omega = -\frac{2\kappa c^2}{H^2 R^2}\left(\frac{\dot{H}}{H} + \frac{\dot{R}}{R}\right) = -\frac{2\kappa c^2}{H^2 R^2}\left[-(1+q)H + H\right] = 2qH(\Omega - 1).$$

The point $\Omega = 1$ is a fixed point of the differential equation. The prefactors H and q determine whether it is a stable or an unstable fixed point. We may assume $H > 0$ for all past times (i.e. the universe was not collapsing in the past). The parameter q is given by

$$q = \frac{4\pi G}{3c^2 H^2}\sum_i \rho_i (1 + 3w_i) = \frac{1}{2}\sum_i \Omega_i (1 + 3w_i).$$

The parameter q is positive if all components i satisfy the strong energy condition $w_i > -\frac{1}{3}$. The notable exception is the vacuum energy, for which we have $w_\Lambda = -1$.

For a radiation dominated universe or a matter dominated universe we have $H > 0$ and $q > 0$. This implies that in these cases the value $\Omega = 1$ is an unstable fixed point.

In a vacuum dominated universe (like ours today) we have $w_\Lambda = -1$ and therefore $H > 0$ and $q < 0$. This implies that for a vacuum dominated universe the value $\Omega = 1$ is a stable fixed point. Note that on cosmological time scales the period where the universe is vacuum dominated is quite recent.

Changing the evolution variable from the time t to the scale factor R, the above differential equation may re-written as

$$\frac{d\Omega}{d\ln R} = 2q\left(\Omega - 1\right).$$

If we assume in addition, that the universe consists only of one component we have

$$q = \frac{1}{2}\Omega\left(1+3w\right)$$

and

$$\frac{d\Omega}{d\ln R} = \left(1+3w\right)\Omega\left(\Omega-1\right).$$

The flatness problem is the following: Given that we measure today (at time t_2) a value of Ω close to 1, we may first evolve back to the time t_1, where the parameter q changed sign. At time t_1 the deviation of Ω from 1 was larger, but still quite close to 1. Evolving backwards would require extremely fine-tuned initial conditions to arrive at this value.

Let us estimate the amplification due to the evolution from t_2 backwards to t_1. We model our universe as consisting of vacuum energy only. We find

$$\Omega_1 - 1 \approx \left(\Omega_2 - 1\right) e^{2H(t_2-t_1)}.$$

With $t_2 - t_1 \approx 13 \cdot 10^9$ yr we find

$$e^{2H(t_2-t_1)} \approx 6.07,$$

and hence

$$\mathcal{O}\left(|\Omega_1 - 1|\right) = \mathcal{O}\left(|\Omega_2 - 1|\right).$$

Let us now consider the backward evolution from t_1 to an earlier time t_0. For simplicity we assume a radiation dominated universe. We now obtain

$$\Omega_0 - 1 \approx \left(\Omega_1 - 1\right)\left(\frac{R_0}{R_1}\right)^2 = \left(\Omega_1 - 1\right)\left(\frac{t_0}{t_1}\right).$$

Taking for t_0 the Planck time $t_0 = 10^{-43}$ s gives

$$\frac{t_0}{t_1} \approx 10^{-61}.$$

11.3 Basics of inflation

Please note that the horizon problem and the flatness problem do not contradict the standard Friedmann-Robertson-Walker cosmology model. All observed phenomena are in agreement with the Friedmann-Robertson-Walker cosmology model and specific initial conditions. The problem is only that the initial conditions need to be extremely fine-tuned to arrive at the universe observed today. We would prefer a mechanism, which starts from rather random initial conditions and nevertheless explains the observations today. This is the motivation for inflationary models.

We call

$$\frac{c}{RH} = \frac{c}{\dot{R}}$$

the **comoving Hubble radius**. During the radiation or matter dominated period, the expansion of the universe decelerates and the comoving Hubble radius increases. The basic idea of inflation is a **shrinking comoving Hubble radius** at the beginning of the universe.

$$\frac{d}{dt}\left(\frac{c}{RH}\right) < 0.$$

This is equivalent to

$$\ddot{R} > 0,$$

or

$$q = -\frac{R\ddot{R}}{\dot{R}^2} < 0.$$

The conditions $\ddot{R} > 0$ or $q < 0$ describe an **accelerated expansion**, hence the name “inflation”.

We may also translate the condition of a shrinking comoving Hubble radius to a condition on the equation of state for a perfect fluid: From the second Friedmann equation

$$\frac{\ddot{R}}{R} = -\frac{4\pi G}{3c^2}(\rho + 3p)$$

we obtain with $\ddot{R} > 0$ the condition

$$p < -\frac{1}{3}\rho.$$

Thus we see that during inflation we had **negative pressure**.

Remark: This violates the strong energy condition ($\rho + 3p \geq 0$), but so does a universe dominated by vacuum energy with an equation of state $p = -\rho$. There is nothing wrong with that, it only means that it is not sensible to impose the strong energy condition.

Let us now discuss how inflation solves the horizon problem and the flatness problem. We first consider the horizon problem. Here, the problem was the finite particle horizon at recombination time. Let us now denote by t_0 the time when inflation ends. During inflation we have an equation of state with $w < -\frac{1}{3}$ or $n < 2$, where $n = 3(1+w)$. The comoving particle horizon at time t_0 is given by

$$\chi_p = \frac{c}{R_2 H_2} \int\limits_{z_0}^{\infty} \frac{dz}{E(z)}.$$

If we assume

$$E(z) = \Omega_2^{\frac{1}{2}} (1+z)^{\frac{n}{2}}$$

with $n < 2$ we obtain for the red shift integral

$$\int\limits_{z_0}^{\infty} dz (1+z)^{-\frac{n}{2}} = \frac{2}{2-n} (1+z)^{\frac{2-n}{2}} \Big|_{z_1}^{\infty}.$$

For $n < 2$ the integral diverges and we obtain an infinite comoving particle horizon. This solves the horizon problem.

Let us now discuss the flatness problem. We recall the differential equation

$$\frac{d\Omega}{d \ln R} = 2q (\Omega - 1).$$

For $q < 0$ the point $\Omega = 1$ is a stable fixed point of the differential equation. Thus, if we start with random initial conditions before inflation, a sufficient long inflation period will evolve the value of Ω very close to 1 at the end of inflation, such that the further evolution according to standard Friedmann-Robertson-Walker cosmology is compatible with the observed value of Ω today. We may estimate the required time period of inflation. Let us denote by t_{-1} the time when inflation starts and by t_0 the time when inflation ends. We assume that

$$|\Omega_{-1} - 1| = \mathcal{O}(1),$$

and

$$|\Omega_0 - 1| = \mathcal{O}\left(10^{-63}\right).$$

It is common practice to give the time of the inflation period by a number N, which corresponds to the power of e, by which the scale factor increases during the inflation period. N is also called the **number of e-folds**. In detail, dN is defined by

$$dN = H dt = d \ln R.$$

Integration yields (with $N(t_{-1}) = 0$ and $N(t_0) = N$)

$$N = \ln \frac{R_0}{R_{-1}} \quad \text{or} \quad \frac{R_0}{R_{-1}} = e^N.$$

Integration of the differential equation

$$\frac{d\Omega}{d\ln R} = (1+3w)\,\Omega\,(\Omega-1)$$

gives

$$\frac{|\Omega_0-1|}{|\Omega_{-1}-1|} = \left(\frac{R_0}{R_{-1}}\right)^{1+3w} = e^{(1+3w)N}$$

and therefore

$$N = -\frac{1}{1+3w}\ln\left(\frac{|\Omega_{-1}-1|}{|\Omega_0-1|}\right).$$

For example, for $|\Omega_{-1}-1|/|\Omega_0-1| = 10^{63}$ and $w=-1$ we obtain $N \approx 73$.

For a perfect fluid with $-1 \leq w < -1/3$ we have $0 \leq n < 2$ and

$$H = H_0\left(\frac{R}{R_0}\right)^{-\frac{n}{2}} = H_0 e^{-\frac{n}{2}N}.$$

For

$$0 \leq \frac{n}{2} \ll 1$$

we see that the Hubble parameter does not change much during inflation. Let us introduce two **slow-roll parameters** ε and η defined by

$$\begin{aligned}
\varepsilon &= -\frac{\dot{H}}{H^2} = -\frac{R}{H}\frac{dH}{dR} = -\frac{d\ln H}{dN}, \\
\eta &= -\frac{\ddot{H}}{2\dot{H}H} = \varepsilon - \frac{1}{2\varepsilon}\frac{d\varepsilon}{dN}.
\end{aligned}$$

We recall that q was defined by $\dot{H} = -(1+q)H^2$ and therefore

$$\varepsilon = 1+q.$$

For the perfect fluid discussed above we find

$$\varepsilon = \frac{n}{2}, \qquad \eta = \frac{n}{2}.$$

11.4 The inflaton field

Up to now we discussed inflation as a period where we have (i) a shrinking comoving Hubble radius, (ii) accelerated expansion, (iii) negative pressure $p < -\rho/3$. The three conditions are (under modest assumptions) equivalent. Let us now discuss how inflation can be realised. We already know about one scenario, which has all the three properties listed above: A universe dominated by vacuum energy. However, just simply postulating that the universe was vacuum dominated early on is not what we want: Such a universe will remain vacuum dominated forever. We would like to have an inflation period, which comes to an end, followed by a radiation dominated period, which is then followed by a matter dominated and finally by a vacuum dominated period.

Let us consider the action of a scalar field minimally coupled to gravity. The action reads

$$\begin{aligned} S &= S_{EH} + S_\phi, \\ S_{EH} &= -\frac{c^3}{16\pi G}\int d^4x\,\sqrt{-g}R, \\ S_\phi &= \frac{1}{c}\int d^4x\,\sqrt{-g}\mathscr{L}_\phi, \qquad \mathscr{L}_\phi = \frac{\hbar^2 c}{2} g^{\mu\nu}\left(\partial_\mu\phi\right)\left(\partial_\nu\phi\right) - V(\phi). \end{aligned}$$

We determine the energy-momentum tensor

$$\begin{aligned} T_{\mu\nu} &= \frac{2}{\sqrt{-g}}\frac{\partial\sqrt{-g}\mathscr{L}}{\partial g^{\mu\nu}} = 2\frac{\partial\mathscr{L}}{\partial g^{\mu\nu}} + \frac{2}{\sqrt{-g}}\mathscr{L}\frac{\partial\sqrt{-g}}{\partial g^{\mu\nu}} = 2\frac{\partial\mathscr{L}}{\partial g^{\mu\nu}} - \mathscr{L}g_{\mu\nu} \\ &= \frac{\hbar^2 c}{2}\left[2\left(\partial_\mu\phi\right)\left(\partial_\nu\phi\right) - g_{\mu\nu}\left(\partial_\lambda\phi\right)\left(\partial^\lambda\phi\right)\right] + g_{\mu\nu}V(\phi). \end{aligned}$$

Let us now specialise to the Robertson-Walker metric with coordinates (ct, r, θ, φ) and assume that the field $\phi(x)$ is homogeneous:

$$\phi(x) = \phi(t,\vec{x}) = \phi(t).$$

This implies

$$\partial_r\phi = \partial_\theta\phi = \partial_\varphi\phi = 0.$$

Then

$$\begin{aligned} T_{00} &= \frac{\hbar^2 c}{2}\left(\partial_0\phi\right)^2 + V(\phi), \\ T_{ij} &= -g_{ij}\left[\frac{\hbar^2 c}{2}\left(\partial_0\phi\right)^2 - V(\phi)\right] \end{aligned}$$

and $T_{0i} = 0$. This is the energy-momentum tensor of a perfect fluid with

$$\rho = \frac{\hbar^2 c}{2}\left(\partial_0\phi\right)^2 + V(\phi), \qquad p = \frac{\hbar^2 c}{2}\left(\partial_0\phi\right)^2 - V(\phi).$$

As usual we define the parameter w by $p = w\rho$, this yields

$$w = \frac{\frac{\hbar^2 c}{2}\left(\partial_0\phi\right)^2 - V(\phi)}{\frac{\hbar^2 c}{2}\left(\partial_0\phi\right)^2 + V(\phi)}.$$

For $V(\phi) > \hbar^2 c(\partial_0\phi)^2$ we have

$$w \;<\; -\frac{1}{3}.$$

Thus we have inflation if the potential energy of the field is larger than twice the kinetic energy of the field.

Let us now discuss the equation of motion for the field ϕ. For the Robertson-Walker metric we have $g^{00} = 1$ and

$$\sqrt{-g} \;=\; \frac{R(t)^3 r^2 \sin\theta}{\sqrt{1-\kappa r^2}}.$$

Therefore

$$\partial_0\sqrt{-g} \;=\; \frac{1}{c}\frac{\partial}{\partial t}\sqrt{-g} \;=\; \frac{3}{c}\frac{\dot{R}}{R}\sqrt{-g} \;=\; \frac{3}{c}H\sqrt{-g}.$$

With the assumptions as above $\mathscr{L}_\phi$ simplifies to

$$\mathscr{L}_\phi \;=\; \frac{\hbar^2 c}{2}\left(\partial_0\phi\right)^2 - V\left(\phi\right)$$

The variation of S_ϕ with respect to the field ϕ gives

$$\begin{aligned}
\delta S_\phi &= \frac{1}{c}\int d^4x\,\sqrt{-g}\left[\hbar^2 c\left(\partial_0\phi\right)\left(\partial_0\delta\phi\right) - \frac{\partial V}{\partial\phi}\delta\phi\right] \\
&= -\frac{1}{c}\int d^4x\,\sqrt{-g}\left[\frac{\hbar^2 c}{\sqrt{-g}}\partial_0\left(\sqrt{-g}\partial_0\phi\right) + \frac{\partial V}{\partial\phi}\right]\delta\phi \\
&= -\frac{1}{c}\int d^4x\,\sqrt{-g}\left[\hbar^2 c\partial_0^2\phi + 3\hbar^2 H\partial_0\phi + \frac{\partial V}{\partial\phi}\right]\delta\phi.
\end{aligned}$$

Thus the equation of motion for the field ϕ is

$$\begin{aligned}
\hbar^2 c\partial_0^2\phi + 3\hbar^2 H\partial_0\phi + \frac{\partial V}{\partial\phi} &= 0. \\
\ddot{\phi} + 3H\dot{\phi} + \frac{c}{\hbar^2}\frac{\partial V}{\partial\phi} &= 0.
\end{aligned}$$

In addition, we have the Friedmann equations, which for $\kappa = 0$ read

$$\begin{aligned}
H^2 &= \frac{8\pi G}{3c^2}\left[\frac{\hbar^2 c}{2}\left(\partial_0\phi\right)^2 + V\left(\phi\right)\right], \\
\frac{\ddot{R}}{R} &= -\frac{8\pi G}{3c^2}\left[\hbar^2 c\left(\partial_0\phi\right)^2 - V\left(\phi\right)\right].
\end{aligned}$$

Previously we introduced the slow-roll parameter $\varepsilon = -\dot{H}/H^2$. We have

$$H^2(1-\varepsilon) = H^2\left(1+\frac{\dot{H}}{H^2}\right) = H^2+\dot{H} = \left(\frac{\dot{R}}{R}\right)^2 + \frac{d}{dt}\left(\frac{\dot{R}}{R}\right) = \frac{\ddot{R}}{R},$$

and therefore

$$\varepsilon = 3\frac{\frac{\hbar^2 c}{2}(\partial_0\phi)^2}{\frac{\hbar^2 c}{2}(\partial_0\phi)^2 + V(\phi)}.$$

Inflation occurs for $\varepsilon \ll 1$, i.e. when the potential energy of the inflaton field dominates over the kinetic energy of the inflaton field. This motivates the name "slow-roll parameter". Inflation ends when $w = -1/3$ or $V(\phi) = \hbar^2 c(\partial_0\phi)^2$. In terms of ε this translates to

$$\varepsilon = 1.$$

The inflation period should be sufficiently long, i.e. $\dot{\phi}$ should not change too fast. Therefore we require

$$\left|\ddot{\phi}\right| \ll \left|3H\dot{\phi}\right|, \left|\frac{c}{\hbar^2}\partial_\phi V\right|.$$

In this limit the equation of motion for the field ϕ simplifies to

$$3H\dot{\phi} + \frac{c}{\hbar^2}\partial_\phi V = 0,$$

and the Friedmann equation to

$$H^2 = \frac{8\pi G}{3c^2}V.$$

In the limit $\varepsilon \ll 1$ the two slow-roll parameters are given by

$$\varepsilon \approx 3\frac{\frac{\hbar^2 c}{2}(\partial_0\phi)^2}{V}, \qquad \eta \approx -\frac{\ddot{\phi}}{H\dot{\phi}}.$$

Thus we see that $|\ddot{\phi}| \ll |3H\dot{\phi}|$ implies $|\eta| \ll 1$. In the limit $\varepsilon \ll 1$ and $|\eta| \ll 1$ we further have

$$\begin{aligned}
\varepsilon &\approx \frac{c^3}{16\pi\hbar^2 G}\left(\frac{\partial_\phi V}{V}\right)^2 = \frac{cM_{Pl}^2}{16\pi\hbar^4}\left(\frac{\partial_\phi V}{V}\right)^2, \\
\eta &\approx \frac{c^3}{8\pi\hbar^2 G}\left(\frac{\partial_\phi^2 V}{V}\right) = \frac{cM_{Pl}^2}{8\pi\hbar^4}\left(\frac{\partial_\phi^2 V}{V}\right).
\end{aligned}$$

Let us now estimate the number of e-folds:

$$\begin{aligned}
N &= \int_{t_i}^{t_f} dN = \int_{t_i}^{t_f} H dt = \int_{\phi_i}^{\phi_f} \frac{H}{\dot{\phi}} d\phi = -\frac{3\hbar^2}{c}\int_{\phi_i}^{\phi_f} \frac{H^2}{\partial_\phi V} d\phi \\
&= -\frac{8\pi G\hbar^2}{c^3}\int_{\phi_i}^{\phi_f} \frac{V}{\partial_\phi V} d\phi = -\frac{8\pi\hbar^4}{cM_{Pl}^2}\int_{\phi_i}^{\phi_f} \frac{V}{\partial_\phi V} d\phi.
\end{aligned}$$

Let us now specialise to the simplest potential

$$V = \frac{1}{2}m^2c^3\phi^2.$$

We have

$$\frac{\partial_\phi V}{V} = \frac{2}{\phi},$$

and

$$\varepsilon = \frac{cM_{Pl}^2}{16\pi\hbar^4}\left(\frac{\partial_\phi V}{V}\right)^2 = \frac{c}{4\pi\hbar^4}\left(\frac{M_{Pl}}{\phi}\right)^2.$$

ϕ_f is determined by $\varepsilon = 1$. We find

$$\phi_f = \frac{1}{\hbar^2}\sqrt{\frac{c}{4\pi}}M_{Pl}.$$

The number of e-folds is

$$N = -\frac{8\pi\hbar^4}{cM_{Pl}^2}\int_{\phi_i}^{\phi_f}\frac{V}{\partial_\phi V}d\phi = -\frac{4\pi\hbar^4}{cM_{Pl}^2}\int_{\phi_i}^{\phi_f}\phi d\phi = -\frac{2\pi\hbar^4}{cM_{Pl}^2}\left(\phi_f^2 - \phi_i^2\right) = \frac{2\pi\hbar^4\phi_i^2}{cM_{Pl}^2} - \frac{1}{2}.$$

The number of e-folds determines how close the density parameter Ω is driven to one during inflation. To solve the flatness problem we require

$$N \gtrsim 60.$$

Remark: The values of the inflaton field in this model are of the order of the Planck mass.

12 Gravitational waves

The first experimental evidence for gravitational waves came from an indirect measurement: the observation of binaries of neutron stars. As the two stars inspiral towards each other they emit gravitational waves. The emission of the gravitational waves is strong enough that it affects the dynamics of the binary system. It carries away energy and angular momentum from the system, reducing the size of the orbit. This effect may occur on a timescale short enough to be observable. This effect has been observed by Hulse and Taylor in 1974 in a binary system consisting of a pulsar and a companion neutron star (Nobel prize 1993).

In 2015 there has been the first direct detection of gravitational waves by the LIGO interferometers, followed by further detections of gravitational waves by the LIGO and VIRGO collaborations (Nobel prize 2017).

12.1 Gauge invariance of gravity

The Einstein-Hilbert action is invariant under **general coordinate transformations**

$$x'^{\mu} = f^{\mu}(x).$$

In fact, one of Einstein's original motivations was to find a theory invariant under these transformations. We may view a general coordinate transformations as a (generalised) gauge transformations. We write an infinitesimal general coordinate transformation as

$$x'^{\mu} = x^{\mu} - \varepsilon \xi^{\mu}(x).$$

The minus sign has no particular importance and is just a convention. The infinitesimal inverse transformation is given by

$$x^{\mu} = x'^{\mu} + \varepsilon \xi^{\mu}(x') + \mathcal{O}(\varepsilon^2).$$

Let us now work out the metric in the transformed system:

$$\begin{aligned}
g'_{\mu'\nu'}(x') &= \frac{\partial x^{\mu}}{\partial x'^{\mu'}} \frac{\partial x^{\nu}}{\partial x'^{\nu'}} g_{\mu\nu}(x(x')) \\
&= \left(\delta^{\mu}_{\mu'} + \varepsilon \partial_{\mu'} \xi^{\mu}(x')\right)\left(\delta^{\nu}_{\nu'} + \varepsilon \partial_{\nu'} \xi^{\nu}(x')\right)\left(g_{\mu\nu}(x') + \varepsilon \xi^{\rho}(x') \partial_{\rho} g_{\mu\nu}(x')\right) + \mathcal{O}(\varepsilon^2) \\
&= g_{\mu'\nu'}(x') + \varepsilon\left[\left(\partial_{\mu'}\xi^{\mu}(x')\right) g_{\mu\nu'}(x') + \left(\partial_{\nu'}\xi^{\nu}(x')\right) g_{\mu'\nu}(x') + \xi^{\rho}(x') \partial_{\rho} g_{\mu'\nu'}(x')\right] \\
&\quad + \mathcal{O}(\varepsilon^2).
\end{aligned}$$

We may write this in a shortened form as

$$g'_{\mu\nu} = g_{\mu\nu} + \varepsilon\left[\left(\partial_{\mu}\xi^{\rho}\right) g_{\rho\nu} + \left(\partial_{\nu}\xi^{\rho}\right) g_{\mu\rho} + \xi^{\rho}\partial_{\rho} g_{\mu\nu}\right] + \mathcal{O}(\varepsilon^2).$$

Let us now specialise to an expansion around the flat Minkowski metric. With

$$g_{\mu\nu}(x) = \eta_{\mu\nu} + \kappa h_{\mu\nu}(x),$$

we find for $h'_{\mu\nu}$:

$$\begin{aligned} h'_{\mu\nu} &= h_{\mu\nu} + \frac{\varepsilon}{\kappa}\left[\left(\partial_\mu \xi^\rho\right)\eta_{\rho\nu} + \left(\partial_\nu \xi^\rho\right)\eta_{\mu\rho}\right] \\ &\quad + \varepsilon\left[\left(\partial_\mu \xi^\rho\right)h_{\rho\nu} + \left(\partial_\nu \xi^\rho\right)h_{\mu\rho} + \xi^\rho \partial_\rho h_{\mu\nu}\right] + \mathcal{O}\left(\varepsilon^2\right). \end{aligned}$$

This expression can be simplified and we find

$$h'_{\mu\nu} = h_{\mu\nu} + \frac{\varepsilon}{\kappa}\left[\nabla_\mu \xi_\nu + \nabla_\nu \xi_\mu\right] + \mathcal{O}\left(\varepsilon^2\right),$$

where $\xi_\mu = g_{\mu\nu}\xi^\nu = \eta_{\mu\nu}\xi^\nu + \kappa h_{\mu\nu}\xi^\nu$. We may view the transformation from $h_{\mu\nu}$ to $h'_{\mu\nu}$ as an **infinitesimal gauge transformation**.

12.2 Linearised gravity

Einstein's equations are non-linear differential equations in the metric. We recall that the Newtonian limit is defined as the limit where

- the gravitational field is weak, such that it can be treated as a perturbation of flat space-time,
- all particle velocities are small compared to the speed of light,
- the gravitational field is static (i.e. time-independent).

In this chapter we are interested in a less restrictive scenario: We consider the situation where the gravitational field is weak, but we will not require that the particle velocities are small nor that the gravitational field is static. (It is clear that we have to allow time-dependent fields in order to describe gravitational waves.) For a weak gravitational field we expand around the flat Minkowski metric

$$g_{\mu\nu} = \eta_{\mu\nu} + \kappa h_{\mu\nu},$$

with

$$\left|\kappa h_{\mu\nu}\right| \ll 1, \quad \mu,\nu \in \{0,1,2,3\}.$$

For $|h_{\mu\nu}| = \mathcal{O}(1)$ this implies

$$\kappa \ll 1,$$

and we may use κ for power counting in perturbation theory. In linearised gravity we keep only the first non-trivial order in an expansion in κ. Since $g_{\mu\nu}$ and $\eta_{\mu\nu}$ are symmetric, $h_{\mu\nu}$ is symmetric as well:

$$h_{\nu\mu} = h_{\mu\nu}.$$

In writing $g_{\mu\nu} = \eta_{\mu\nu} + \kappa h_{\mu\nu}$ we have picked a reference frame and broken the invariance under general coordinate transformations. However, there remains a residual freedom in the choice of coordinates. Under an infinitesimal transformation

$$x'^{\mu} = x^{\mu} - \varepsilon \xi^{\mu}(x).$$

we have

$$\begin{aligned}
\kappa h'_{\mu\nu} &= \kappa h_{\mu\nu} + \varepsilon \left[\left(\partial_\mu \xi^\rho\right) \eta_{\rho\nu} + \left(\partial_\nu \xi^\rho\right) \eta_{\mu\rho} \right] + \mathcal{O}\left(\varepsilon^2, \varepsilon\kappa\right) \\
&= \kappa h_{\mu\nu} + \varepsilon \left(\partial_\mu \xi_\nu + \partial_\nu \xi_\mu \right) + \mathcal{O}\left(\varepsilon^2, \varepsilon\kappa\right).
\end{aligned}$$

To lowest order we have $\xi_\mu = g_{\mu\nu}\xi^\nu = \eta_{\mu\nu}\xi^\nu + \mathcal{O}(\kappa)$. We have $|\kappa h'_{\mu\nu}| \ll 1$ provided $\mathcal{O}(\varepsilon) = \mathcal{O}(\kappa)$ and

$$\left| \partial_\mu \xi_\nu \right| = \mathcal{O}(1),$$

i.e. the coordinate transformation is not fastly varying. We call these coordinate transformations **gauge transformations in the linearised theory**.

Let us now work out the expressions for the most important quantities in linearised gravity. The inverse metric is given by

$$g^{\mu\nu} = \eta^{\mu\nu} - \kappa h^{\mu\nu} + \mathcal{O}\left(\kappa^2\right),$$

where $h^{\mu\nu}$ is given by

$$h^{\mu\nu} = \eta^{\mu\rho}\eta^{\nu\sigma} h_{\rho\sigma}.$$

In general we may rise and lower indices with $\eta^{\mu\nu}$ and $\eta_{\mu\nu}$ in quantities which are first order in κ. The Christoffel symbols in linearised gravity are

$$\begin{aligned}
\Gamma^{\rho}{}_{\mu\nu} &= \frac{1}{2} g^{\rho\lambda} \left(\partial_\mu g_{\nu\lambda} + \partial_\nu g_{\mu\lambda} - \partial_\lambda g_{\mu\nu} \right) \\
&= \frac{\kappa}{2} \eta^{\rho\lambda} \left(\partial_\mu h_{\nu\lambda} + \partial_\nu h_{\mu\lambda} - \partial_\lambda h_{\mu\nu} \right) + \mathcal{O}\left(\kappa^2\right) \\
&= \frac{\kappa}{2} \left(\partial_\mu h_\nu{}^\rho + \partial_\nu h_\mu{}^\rho - \partial^\rho h_{\mu\nu} \right) + \mathcal{O}\left(\kappa^2\right).
\end{aligned}$$

Since the Christoffel symbols are first order in κ, we need to keep for the Riemann curvature tensor only the derivatives of the Christoffel symbols, but not the Γ^2-terms:

$$\begin{aligned}
R^{\rho}{}_{\sigma\mu\nu} &= \partial_\mu \Gamma^{\rho}_{\nu\sigma} - \partial_\nu \Gamma^{\rho}_{\mu\sigma} + \Gamma^{\eta}_{\nu\sigma}\Gamma^{\rho}_{\mu\eta} - \Gamma^{\eta}_{\mu\sigma}\Gamma^{\rho}_{\nu\eta} \\
&= \partial_\mu \Gamma^{\rho}_{\nu\sigma} - \partial_\nu \Gamma^{\rho}_{\mu\sigma} + \mathcal{O}\left(\kappa^2\right) \\
&= \frac{\kappa}{2} \left(\partial_\mu \partial_\sigma h_\nu{}^\rho - \partial_\nu \partial_\sigma h_\mu{}^\rho - \partial_\mu \partial^\rho h_{\nu\sigma} + \partial_\nu \partial^\rho h_{\mu\sigma} \right) + \mathcal{O}\left(\kappa^2\right).
\end{aligned}$$

The Ricci tensor is given by

$$\begin{aligned}
Ric_{\mu\nu} &= R^{\lambda}{}_{\mu\lambda\nu} \\
&= \frac{\kappa}{2} \left(\partial_\mu \partial_\rho h_\nu{}^\rho + \partial_\nu \partial_\rho h_\mu{}^\rho - \partial_\mu \partial_\nu h_\rho{}^\rho - \Box h_{\mu\nu} \right) + \mathcal{O}\left(\kappa^2\right).
\end{aligned}$$

For the scalar curvature we obtain

$$R = \kappa\left(\partial^\mu\partial^\nu h_{\mu\nu} - \Box h_{\mu\nu}\eta^{\mu\nu}\right) + \mathcal{O}\left(\kappa^2\right).$$

Putting everything together we obtain for the Einstein tensor

$$\begin{aligned} G_{\mu\nu} &= Ric_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R \\ &= \frac{\kappa}{2}\left(\partial_\mu\partial_\rho h_\nu{}^\rho + \partial_\nu\partial_\rho h_\mu{}^\rho - \partial_\mu\partial_\nu h_\rho{}^\rho - \Box h_{\mu\nu} - \eta_{\mu\nu}\partial^\rho\partial^\sigma h_{\rho\sigma} + \eta_{\mu\nu}\Box h_{\rho\sigma}\eta^{\rho\sigma}\right) + \mathcal{O}\left(\kappa^2\right). \end{aligned}$$

Let us introduce the **trace-reversed perturbation**:

$$\bar{h}_{\mu\nu} = h_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}h, \quad h = h_{\rho\sigma}\eta^{\rho\sigma}.$$

We have

$$\bar{h} = \eta^{\mu\nu}\bar{h}_{\mu\nu} = h - \frac{1}{2}\eta^{\mu\nu}\eta_{\mu\nu}h = -h.$$

This motivates the name "trace-reversed perturbation". The inverse transformation from $\bar{h}_{\mu\nu}$ to $h_{\mu\nu}$ is given by

$$h_{\mu\nu} = \bar{h}_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}\bar{h}.$$

In terms of $\bar{h}$ the Einstein tensor reads

$$G_{\mu\nu} = \frac{\kappa}{2}\left(\partial_\mu\partial_\rho\bar{h}_\nu{}^\rho + \partial_\nu\partial_\rho\bar{h}_\mu{}^\rho - \Box\bar{h}_{\mu\nu} - \eta_{\mu\nu}\partial^\rho\partial^\sigma\bar{h}_{\rho\sigma}\right) + \mathcal{O}\left(\kappa^2\right).$$

The expression for the Einstein tensor is slightly simpler when expressed through $\bar{h}$ instead of h. Einstein's equations read now

$$\Box\bar{h}_{\mu\nu} + \eta_{\mu\nu}\partial^\rho\partial^\sigma\bar{h}_{\rho\sigma} - \partial_\mu\partial_\rho\bar{h}_\nu{}^\rho - \partial_\nu\partial_\rho\bar{h}_\mu{}^\rho = -\frac{16\pi G}{\kappa c^4}T_{\mu\nu} + \mathcal{O}(\kappa).$$

This equation simplifies if we choose a coordinate system in which

$$\partial^\nu\bar{h}_{\mu\nu} = 0.$$

This equation defines the **Lorenz gauge**. Due to the freedom of gauge transformations in the linearised theory we may always impose this condition. To see this, assume that $\bar{h}_{\mu\nu}$ is not of this form. Under the gauge transformation $x'^\mu = x^\mu - \kappa\xi^\mu(x)$ we have

$$\bar{h}'_{\mu\nu} = \bar{h}_{\mu\nu} + \left(\partial_\mu\xi_\nu + \partial_\nu\xi_\mu - \eta_{\mu\nu}\partial^\rho\xi_\rho\right) + \mathcal{O}(\kappa),$$

and

$$\partial^\nu\bar{h}'_{\mu\nu} = \partial^\nu\bar{h}_{\mu\nu} + \Box\xi_\mu + \mathcal{O}(\kappa),$$

If we want to enforce $\partial^\nu \bar{h}'_{\mu\nu} = 0$ we have to find a ξ_μ such that

$$\Box \xi_\mu = -\partial^\nu \bar{h}_{\mu\nu}.$$

From the theory of Green functions we know that a solution of the equation

$$\Box \xi_\mu(x) = j_\mu(x)$$

is given by

$$\xi_\mu(x) = \int d^4y\, G(x-y)\, j_\mu(y),$$

where $G(x-y)$ is the Green function satifying

$$\Box_x G(x-y) = \delta^4(x-y).$$

Thus we see that in Lorenz gauge the linearised Einstein's equations take the form of a wave equation with a source (neglecting $\mathcal{O}(\kappa)$-terms):

$$\Box \bar{h}_{\mu\nu} = -\frac{16\pi G}{\kappa c^4} T_{\mu\nu}.$$

Outside the source we have

$$\Box \bar{h}_{\mu\nu} = 0.$$

Let us count the degrees of freedom. We have

$$\begin{aligned} \bar{h}_{\mu\nu} = \bar{h}_{\nu\mu} &\Rightarrow 10\,\text{d.o.f.}, \\ \partial^\nu \bar{h}_{\mu\nu} = 0 &\Rightarrow (10-4)\ \text{d.o.f.} = 6\,\text{d.o.f.}. \end{aligned}$$

However, imposing just the Lorenz gauge does not eliminate completely all gauge freedom. We may still perform gauge transformations

$$x'^\mu = x^\mu - \kappa\xi^\mu, \quad \text{with } \Box\xi^\mu = 0.$$

Under these transformation we have

$$\begin{aligned} \bar{h}'_{\mu\nu} &= \bar{h}_{\mu\nu} + \xi_{\mu\nu} + \mathcal{O}(\kappa), \qquad \xi_{\mu\nu} = \partial_\mu\xi_\nu + \partial_\nu\xi_\mu - \eta_{\mu\nu}\partial^\rho\xi_\rho, \\ \partial^\nu \bar{h}'_{\mu\nu} &= \partial^\nu \bar{h}_{\mu\nu} + \mathcal{O}(\kappa), \end{aligned}$$

and we see that we stay within Lorenz gauge if the original field $h_{\mu\nu}$ satisfies the Lorenz condition. Thus we may impose four additional constraints on $\bar{h}_{\mu\nu}$. A possible choice of additional constraints is

$$\begin{aligned} \bar{h} &= 0, \\ \bar{h}_{0i} &= 0, \quad i \in \{1,2,3\}. \end{aligned}$$

The first additional constraint ($\bar{h} = 0$) implies

$$\bar{h}_{\mu\nu} = h_{\mu\nu},$$

the second additional constraint combined with the Lorenz condition implies

$$\partial^\nu \bar{h}_{0\nu} = \partial^0 \bar{h}_{00} + \partial^i \bar{h}_{0i} = \partial^0 \bar{h}_{00} = \partial^0 h_{00} = 0.$$

Thus h_{00} is a time-independent or non-dynamical component. If non-zero, it corresponds to a static Newtonian potential. For the discussion of gravitational waves we are not interested in static components and one takes $h_{00} = 0$. Technically, we replace the first Lorenz condition $\partial^\nu \bar{h}_{0\nu} = 0$ by $h_{00} = 0$. Thus we arrive at the conditions for the **transverse traceless gauge**:

$$\begin{aligned} h_{0\mu} &= 0, \\ h_i{}^i &= 0, \\ \partial^j h_{ij} &= 0. \end{aligned}$$

One easily checks that these conditions imply the Lorenz condition $\partial^\nu \bar{h}_{\mu\nu} = 0$. It is common practice to denote the field $h_{\mu\nu}$ in the transverse traceless gauge by

$$h^{\mathrm{TT}}_{\mu\nu},$$

where TT stands for "transverse traceless". Let us now count again the degrees of freedom. We have

$$\begin{aligned} h_{\mu\nu} = h_{\nu\mu} &\Rightarrow 10 \text{ d.o.f.}, \\ h_{0\mu} = 0 &\Rightarrow (10-4) \text{ d.o.f.} = 6 \text{ d.o.f.}, \\ h_i{}^i = 0 &\Rightarrow (10-4-1) \text{ d.o.f.} = 5 \text{ d.o.f.}, \\ \partial^j h_{ij} = 0 &\Rightarrow (10-4-1-3) \text{ d.o.f.} = 2 \text{ d.o.f.}. \end{aligned}$$

Thus we are left with 2 independent components of the metric, which correspond to the two physical degrees of freedom for a gravitational wave.

In a particle picture the field $h_{\mu\nu}$ describes a graviton, which is a massless spin-2 particle, whose helicity states are only $+2$ and -2. This is similar to the photon field A_μ, which describes a massless spin-1 particle, whose helicity states are only $+1$ and -1.

Remark: One may choose the transverse traceless gauge in vacuum ($T_{\mu\nu} = 0$), but not inside the source. The transverse traceless gauge imposes $h_{00} = 0$, which implies that there are no static components. This is true in the vacuum (far away from the sources), but not inside the sources. Inside the source we may decompose $h_{\mu\nu}$ into

- unphysical gauge degrees of freedom,
- physical non-radiative degrees of freedom related to matter sources,
- physical radiative degrees of freedom.

By a careful analysis one may show, that the physical radiative degrees of freedom obey a wave equation, while the physical non-radiative degrees of freedom obey a Poisson equation (i.e. an equation of the type $\Delta\phi = -4\pi\rho$). For the piece of the metric corresponding to the physical radiative degrees of freedom one may impose the transverse traceless conditions.

Let us now return to the wave equation in the vacuum:

$$\Box h^{\mathrm{TT}}_{\mu\nu} = 0.$$

Solutions to this equation are

$$h^{\mathrm{TT}}_{\mu\nu} = C_{\mu\nu} e^{\pm i k_\rho x^\rho}, \qquad k_\rho = \left(\frac{\omega}{c}, \vec{k}\right),$$

where $C_{\mu\nu}$ is a constant symmetric rank-2 tensor, which is purely spatial and traceless:

$$\begin{aligned} C_{0\mu} &= 0, \\ C_\mu{}^\mu &= 0. \end{aligned}$$

Note that $C_{00} = 0$ implies $C_\mu{}^\mu = C_j{}^j$. We have

$$\Box h^{\mathrm{TT}}_{\mu\nu} = C_{\mu\nu} \Box e^{\pm i k_\rho x^\rho} = -C_{\mu\nu} k^2 e^{\pm i k_\rho x^\rho}$$

and $C_{\mu\nu} e^{\pm i k_\rho x^\rho}$ is a solution of the wave equation provided $k^2 = 0$. Thus

$$\frac{\omega}{c} = |\vec{k}|.$$

The condition $k^2 = 0$ also implies that gravitational waves propagate with the speed of light. We set $k = |\vec{k}|$ and write

$$\vec{k} = k\hat{n},$$

where $\hat{n}$ is a unit vector ($|\hat{n}| = 1$). Of course, the perturbation of the metric should be real. This is easily enforced by replacing $e^{ik\cdot x}$ and $e^{-ik\cdot x}$ by

$$\cos(k\cdot x), \qquad \sin(k\cdot x).$$

Let us investigate the polarisation tensor $C_{\mu\nu}$ in more detail. As $C_{\mu\nu}$ is purely spatial we have

$$C_{\mu\nu} = \left(\begin{array}{cccc} 0 & 0 & 0 & 0 \\ 0 & & & \\ 0 & & C_{ij} & \\ 0 & & & \end{array} \right).$$

The Lorenz condition implies

$$\partial^j h_{ij} = 0 \;\Rightarrow\; \hat{n}^j C_{ij} = 0.$$

$\hat{n}$ is a vector in the three-dimensional spatial sub-space and defines a plane in this sub-space as the vectors perpendicular to $\hat{n}$. Let $\hat{u}$ and $\hat{v}$ be two orthogonal unit vectors in this plane. With the help of $\hat{u}$ and $\hat{v}$ we may express the two polarisation states of C_{ij} as

$$\begin{aligned} C^{+}_{ij} &= \hat{u}_i \hat{u}_j - \hat{v}_i \hat{v}_j, \\ C^{\times}_{ij} &= \hat{u}_i \hat{v}_j + \hat{v}_i \hat{u}_j. \end{aligned}$$

One easily verifies that C^+ and $C^\times$ satisfy the traceless condition:

$$\begin{aligned} \left(C^+\right)_j{}^j &= -\hat{u}^2 + \hat{v}^2 = -1 + 1 = 0, \\ \left(C^\times\right)_j{}^j &= \hat{u}_j \hat{v}^j + \hat{v}_j \hat{u}^j = -\hat{u}\cdot\hat{v} - \hat{v}\cdot\hat{u} = -2\hat{u}\cdot\hat{v} = 0. \end{aligned}$$

Let us now specialise to $\hat{n} = \hat{e}_z$and $\hat{u} = \hat{e}_x$, $\hat{v} = \hat{e}_y$. We define

$$h_+ = C_{11}, \qquad h_\times = C_{12}.$$

Then

$$h^{\mathrm{TT}}_{\mu\nu} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & h_+ & h_\times & 0 \\ 0 & h_\times & -h_+ & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \cos\left(\omega\left(t - \frac{z}{c}\right)\right).$$

Let $\hat{n}$ be a three-dimensional unit vector

$$\sum_{j=1}^{3} \hat{n}_j \hat{n}^j = 1,$$

describing the propagation direction of a gravitational wave. Given a solution $\bar{h}_{\mu\nu}$ of $\Box \bar{h}_{\mu\nu} = 0$ in Lorenz gauge, we may easily project on the transverse traceless gauge as follows: We first define projection operators

$$P_i{}^j = \delta_i{}^j - \hat{n}_i \hat{n}^j, \quad P_{ij} = P_i{}^k \eta_{kj} = \eta_{ij} - \hat{n}_i \hat{n}_j, \quad \Lambda_{ij}{}^{kl} = P_i{}^k P_j{}^l - \frac{1}{2} P_{ij} P^{kl}.$$

We use the convention that repeated latin indices $i, j, ...$ are summed over $\{1,2,3\}$. $\Lambda_{ij}{}^{kl}$ satisfies

$$\begin{aligned} &\Lambda_{ij}{}^{kl} \hat{n}^i = \Lambda_{ij}{}^{kl} \hat{n}^j = \Lambda_{ij}{}^{kl} \hat{n}_k = \Lambda_{ij}{}^{kl} \hat{n}_l = 0, \\ &\Lambda_{ij}{}^{kl} \Lambda_{kl}{}^{nm} = \Lambda_{ij}{}^{nm}, \\ &\eta^{ij} \Lambda_{ij}{}^{kl} = \Lambda_{ij}{}^{kl} \eta_{kl} = 0. \end{aligned}$$

We then define $h^{\mathrm{TT}}_{\mu\nu}$ by $h^{\mathrm{TT}}_{0\mu} = 0$ and

$$h^{\mathrm{TT}}_{ij} = \Lambda_{ij}{}^{kl} \bar{h}_{kl}.$$

One easily verifies that $h^{\mathrm{TT}}_{\mu\nu}$ is in the transverse traceless gauge:

$$\begin{aligned}
\eta^{ij} h^{\mathrm{TT}}_{ij} &= \eta^{ij} \Lambda_{ij}{}^{kl} \bar{h}_{kl} \;=\; 0, \\
\partial^j h^{\mathrm{TT}}_{ij} &= \pm i k \hat{n}^j h^{\mathrm{TT}}_{ij} \;=\; \pm i k \hat{n}^j \Lambda_{ij}{}^{kl} \bar{h}_{kl} \;=\; 0.
\end{aligned}$$

Remark: The polarisation states of a classical radiation field are related to the particles that one obtains upon quantisation. In particular one obtains the spin of the quantised field from the transformation properties of the polarisation modes: If the polarisation modes are invariant under a rotation of an angle θ, the spin of the quantised particle is given by

$$S = \frac{2\pi}{\theta}.$$

Let us consider a rotation in the x-y-plane:

$$R_{ij} = \begin{pmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

The field transforms as

$$\left(h^{\mathrm{TT}}_{ij}\right)' = R_{ik} R_{jl} h^{\mathrm{TT}}_{kl}.$$

Explicitly we find

$$\begin{aligned}
h'_{+} &= h_{+} \cos(2\theta) + h_{\times} \sin(2\theta), \\
h'_{\times} &= -h_{+} \sin(2\theta) + h_{\times} \cos(2\theta).
\end{aligned}$$

This is invariant for $\theta = \pi$ and therefore

$$S = \frac{2\pi}{\pi} = 2.$$

12.3 Detection of gravitational waves

In this section we investigate the effect of gravitational waves on test masses. The metric of a gravitational wave is given by

$$g_{\mu\nu} = \eta_{\mu\nu} + \kappa h^{\mathrm{TT}}_{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} + \kappa \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & h_{+} & h_{\times} & 0 \\ 0 & h_{\times} & -h_{+} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \cos\left(\omega\left(t - \frac{z}{c}\right)\right).$$

In general relativity, a particle moves along a geodesic. The geodesic equation reads

$$\frac{d^2 x^\mu}{ds^2} + \Gamma^\mu_{\tau\sigma} \frac{dx^\tau}{ds} \frac{dx^\sigma}{ds} = 0,$$

where $s = c\tau$ and τ is the proper time. Let us work out the coordinate acceleration d^2x^i/dt^2 for $i \in \{1,2,3\}$:

$$\begin{aligned}
\frac{d^2x^i}{c^2dt^2} &= \frac{d^2x^i}{d(x^0)^2} = \left(\frac{ds}{dx^0}\right)\frac{d}{ds}\left[\frac{dx^i}{ds}\frac{ds}{dx^0}\right] \\
&= \left(\frac{dx^0}{ds}\right)^{-2}\frac{d^2x^i}{ds^2} - \left(\frac{dx^0}{ds}\right)^{-3}\left(\frac{dx^i}{ds}\right)\frac{d^2x^0}{ds^2} \\
&= -\left(\frac{dx^0}{ds}\right)^{-2}\Gamma^i{}_{\mu\nu}\frac{dx^\mu}{ds}\frac{dx^\nu}{ds} + \left(\frac{dx^0}{ds}\right)^{-3}\left(\frac{dx^i}{ds}\right)\Gamma^0{}_{\mu\nu}\frac{dx^\mu}{ds}\frac{dx^\nu}{ds} \\
&= -\left(\frac{dx^0}{ds}\right)^{-2}\left[\Gamma^i{}_{00}\frac{dx^0}{ds}\frac{dx^0}{ds} + 2\Gamma^i{}_{0j}\frac{dx^0}{ds}\frac{dx^j}{ds} + \Gamma^i{}_{jk}\frac{dx^j}{ds}\frac{dx^k}{ds}\right] \\
&\quad + \left(\frac{dx^0}{ds}\right)^{-3}\left(\frac{dx^i}{ds}\right)\left[\Gamma^0{}_{00}\frac{dx^0}{ds}\frac{dx^0}{ds} + 2\Gamma^0{}_{0j}\frac{dx^0}{ds}\frac{dx^j}{ds} + \Gamma^0{}_{jk}\frac{dx^j}{ds}\frac{dx^k}{ds}\right] \\
&= -\Gamma^i{}_{00} - 2\Gamma^i{}_{0j}\frac{dx^j}{dx^0} - \Gamma^i{}_{jk}\frac{dx^j}{dx^0}\frac{dx^k}{dx^0} + \frac{dx^i}{dx^0}\left(\Gamma^0{}_{00} + 2\Gamma^0{}_{0j}\frac{dx^j}{dx^0} + \Gamma^0{}_{jk}\frac{dx^j}{dx^0}\frac{dx^k}{dx^0}\right) \\
&= -\Gamma^i{}_{00} - 2\Gamma^i{}_{0j}\frac{v^j}{c} - \Gamma^i{}_{jk}\frac{v^jv^k}{c^2} + \frac{v^i}{c}\left(\Gamma^0{}_{00} + 2\Gamma^0{}_{0j}\frac{v^j}{c} + \Gamma^0{}_{jk}\frac{v^jv^k}{c^2}\right).
\end{aligned}$$

In the last line we introduced the coordinate velocities $v^i = dx^i/dt$. Let us assume that our test mass is initially at rest in our coordinate system. In this case the geodesic equation reduces to

$$\frac{d^2x^i}{c^2dt^2} + \Gamma^i{}_{00} = 0.$$

In the transverse traceless gauge we have

$$\Gamma^i{}_{00} = \frac{\kappa}{2}\left(\partial_0 h_0{}^i + \partial_0 h_0{}^i - \partial^i h_{00}\right) = 0$$

and hence

$$\frac{d^2x^i}{c^2dt^2} = 0.$$

Thus

$$x^i = \text{const.}$$

This does not mean that a gravitational wave has no effect on test masses. It only means that our chosen coordinate system moves with the waves.

In order to understand the situation consider a spherical balloon, where we mark a few points. As coordinates on the surface of the balloon we use two angles θ and ϕ. We then periodically increase/decrease the air inside the balloon. Thus the metric on the surface of the balloon is given by

$$ds^2 = \left[R_0 + A\cos(\omega t)\right]^2\left[d\theta^2 + \sin^2\theta d\phi^2\right].$$

The marked points stay at constant θ and ϕ, however the distance between two marked points is varying with time.

Let us now return to gravitational waves. We have to look at the distance between two test masses. Consider a gravitational wave propagating along the z-direction. Assume that at each of the two points

$$P_1 : (x,y,z) \;=\; (0,0,0), \qquad P_2 : (x,y,z) \;=\; (x_2,0,0)$$

we have a test mass. The distance between the two points is

$$\begin{aligned}
d_{12} &= \int\limits_0^{x_2} dx \sqrt{|g_{11}|} \;=\; \int\limits_0^{x_2} dx \sqrt{\left|-1+\kappa h_{11}^{\mathrm{TT}}\right|} \;=\; \int\limits_0^{x_2} dx \sqrt{1-\kappa h_+ \cos\left(\omega\left(t-\frac{z}{c}\right)\right)} \\
&= x_2 \sqrt{1-\kappa h_+ \cos\left(\omega\left(t-\frac{z}{c}\right)\right)} \\
&\approx x_2 \left(1-\frac{\kappa h_+}{2} \cos\left(\omega\left(t-\frac{z}{c}\right)\right)\right).
\end{aligned}$$

Thus we see that the distance between the two test masses is changing with time in the presence of a gravitational wave. For the fractional distance change one has

$$\frac{\delta d_{12}}{d_{12}} \;\approx\; -\frac{\kappa h_+}{2} \cos\left(\omega\left(t-\frac{z}{c}\right)\right).$$

Let us now repeat the calculation, where the point P_2 has coordinates

$$P_2 : (x,y,z) \;=\; (\cos\varphi, \sin\varphi, 0).$$

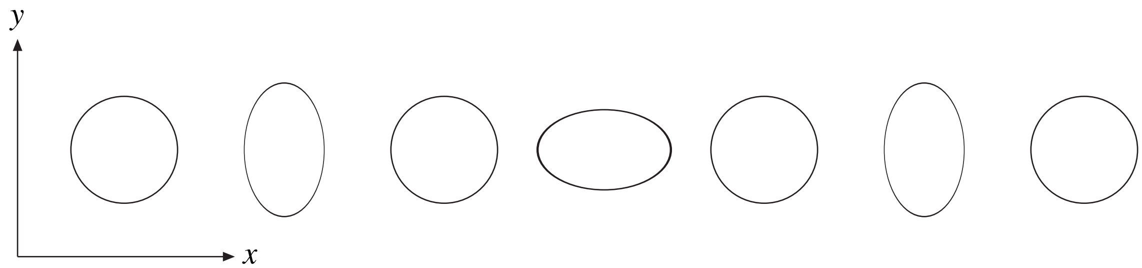
We now have

$$\begin{aligned}
d_{12} &= \int\limits_0^{1} d\lambda \sqrt{\left|g_{11}\cos^2\varphi + g_{22}\sin^2\varphi + 2 g_{12}\sin\varphi\cos\varphi\right|} \\
&= \sqrt{1-\kappa h_+ \cos(2\varphi)\cos\left(\omega\left(t-\frac{z}{c}\right)\right) - \kappa h_\times \sin(2\varphi)\cos\left(\omega\left(t-\frac{z}{c}\right)\right)} \\
&\approx 1-\frac{\kappa h_+}{2}\cos(2\varphi)\cos\left(\omega\left(t-\frac{z}{c}\right)\right) - \frac{\kappa h_\times}{2}\sin(2\varphi)\cos\left(\omega\left(t-\frac{z}{c}\right)\right).
\end{aligned}$$

Let us now specialise to case, where the gravitaional wave has a pure "plus"-polarisation, i.e. $h_+ \neq 0$, $h_\times = 0$. In this case

$$\frac{\delta d_{12}}{d_{12}} \;\approx\; -\frac{\kappa h_+}{2} \cos(2\varphi)\cos\left(\omega\left(t-\frac{z}{c}\right)\right).$$

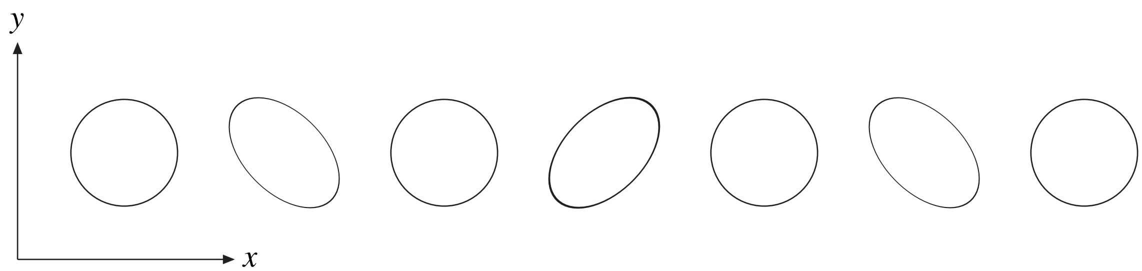
Consider now a test mass at the origin in the $x-y$-plane and a number of test masses on a circle in the $x-y$-plane with centre $(0,0)$. Plotting the distance between the test mass at the centre and a test mass at an angle φ at various times gives us the following picture for the "plus"-polarisation:



For a gravitational wave, which has a pure "cross"-polarisation, i.e. $h_+ = 0$ and $h_\times \neq 0$ we have

$$\frac{\delta d_{12}}{d_{12}} \approx -\frac{\kappa h_\times}{2} \sin(2\varphi) \cos\left(\omega\left(t - \frac{z}{c}\right)\right).$$

This gives us the following picture for the the "cross"-polarisation:

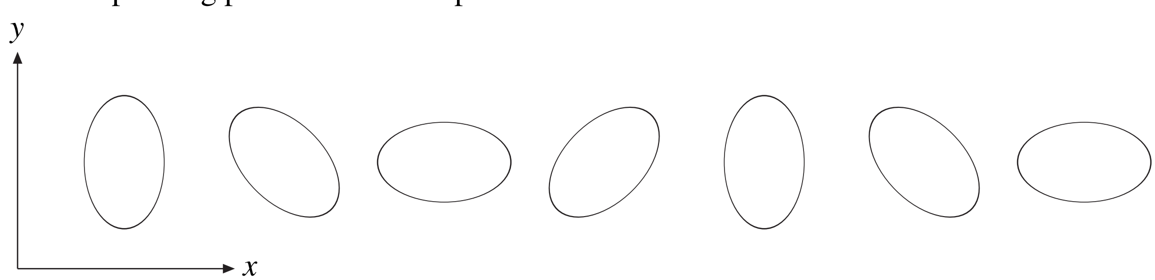


These plots clarify also the motivation for the names "plus"-polarisation and "cross"-polarisation.

As in optics, we may consider linear combinations of the "plus"-polarisation and the "cross"-polarisation. Left- and right-circular polarisations are defined by

$$h_{\mu\nu}^{\mathrm{TT}} = C_{\mu\nu}^{+} \cos\left(\omega\left(t - \frac{z}{c}\right)\right) \pm C_{\mu\nu}^{\times} \sin\left(\omega\left(t - \frac{z}{c}\right)\right).$$

The corresponding plot for a circular polarisation looks as follows:



Typical experiments for the detection of gravitational waves work as follows: A laser beam splitter is placed at

$$P_1 : (x, y, z) = (0, 0, 0).$$

Two mirrors are placed at

$$P_2 : (x, y, z) = (L, 0, 0), \qquad P_3 : (x, y, z) = (0, L, 0).$$

A laser beam is sent to the splitter, the light travels than along the x-arm and y-arm. At the mirrors it is reflected. At the splitter, the two beams are combined again and sent to the detector. A non-equal change in the distance from the splitter to the mirrors will result in an observable interference pattern at the detector.

A typical range for observable gravitational wave frequencies for terrestial detectors is

$$0.1\mathrm{s}^{-1} < \omega < 10^3\mathrm{s}^{-1}.$$

Typical values for the amplitude of the metric perturbation are

$$\left|\kappa h_{\mu\nu}\right| = \mathcal{O}\left(10^{-21}\right).$$

12.4 Production of gravitational waves

Let us now consider the production of gravitational waves. To simplify the derivation we will again assume that the gravitational field is weak. In addition we assume that the velocities of the sources are small. The need for the second assumption can be anticipated from classical Newtonian mechanics. The virial theorem for a two-body system with an $1/r$-potential states

$$\begin{aligned} \langle T \rangle &= -\frac{1}{2}\langle U \rangle, \\ \frac{1}{2}\mu v^2 &= \frac{1}{2}\frac{G\mu M}{r}, \qquad M = m_1 + m_2, \quad \mu = \frac{m_1 m_2}{m_1 + m_2}. \end{aligned}$$

Here, $\langle \dots \rangle$ denotes the time average. We denote the Scharzschild radius by

$$r_s = \frac{2GM}{c^2}.$$

Then

$$\left(\frac{v}{c}\right)^2 = \frac{1}{2}\frac{r_s}{r}.$$

A weak gravitational field implies $r_s \ll r$, within classical Newtonian mechanics this implies $v \ll c$.

Let us now discuss the generation of gravitational waves. We start from Einstein's equations with a source term. In Lorenz gauge we have

$$\Box \bar{h}_{\mu\nu} = -\frac{16\pi G}{\kappa c^4} T_{\mu\nu}.$$

We recall from electrodynamics that an equation of the form

$$\Box f(x) = j(x)$$

is solved with the help of the Green's function. By definition, the Green's function $G(x,x') = G(ct,\vec{x},ct',\vec{x}')$ satisfies

$$\Box G\left(x,x'\right) = \delta^4\left(x-x'\right) = \frac{1}{c}\delta\left(t-t'\right)\delta^3\left(\vec{x}-\vec{x}'\right).$$

The Green's function for the d'Alembert operator

$$\Box = \frac{1}{c^2}\frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x^2} - \frac{\partial^2}{\partial y^2} - \frac{\partial^2}{\partial z^2}$$

is well-known:

$$G^{\pm}(ct,\vec{x},ct',\vec{x}') = \frac{1}{4\pi}\frac{1}{|\vec{x}-\vec{x}'|}\delta\left(ct - \left[ct' \pm |\vec{x}-\vec{x}'|\right]\right).$$

G^+ is called the **retarded Green's function**, G^- is called the **advanced Green's function**. In the following we will only consider the retarded Green's function and drop the superscript "+". A solution to $\Box f(x) = j(x)$ is given by

$$f(x) = \int d^4x' G\left(x,x'\right) j\left(x'\right).$$

Thus

$$\begin{aligned} \bar{h}_{\mu\nu} &= -\frac{16\pi G}{\kappa c^4}\int d^4x' G\left(x,x'\right) T_{\mu\nu}\left(x'\right) \\ &= -\frac{4G}{\kappa c^4}\int d^3x' \frac{T_{\mu\nu}\left(ct - |\vec{x}-\vec{x}'|,\vec{x}'\right)}{|\vec{x}-\vec{x}'|}. \end{aligned}$$

We are in particular interested in the spatial part

$$\bar{h}_{ij} = -\frac{4G}{\kappa c^4}\int d^3x' \frac{T_{ij}\left(ct - |\vec{x}-\vec{x}'|,\vec{x}'\right)}{|\vec{x}-\vec{x}'|}.$$

Far away from the source we may approximate $|\vec{x}-\vec{x}'|$ by

$$\left|\vec{x}-\vec{x}'\right| = r - \hat{n}\cdot\vec{x}' + \mathcal{O}\left(r^{-1}\right), \quad r = |\vec{x}|, \quad \hat{n} = \frac{\vec{x}}{|\vec{x}|} = \frac{\vec{x}}{r}.$$

For the energy-momentum tensor we have

$$T_{ij}\left(ct - \left|\vec{x}-\vec{x}'\right|,\vec{x}'\right) \approx T_{ij}\left(ct - r + \hat{n}\cdot\vec{x}',\vec{x}'\right) \approx T_{ij}\left(ct-r,\vec{x}'\right) + \left(\hat{n}\cdot\vec{x}'\right)\partial_0 T_{ij}\left(ct-r,\vec{x}'\right).$$

Let t_s be the typical time scale of variation of the source, i.e.

$$\partial_0 T_{ij} \approx \frac{T_{ij}}{ct_s}.$$

Then

$$\left(\hat{n}\cdot\vec{x}'\right)\partial_0 T_{ij} \;\approx\; \left(\hat{n}\cdot\vec{x}'\right)\frac{T_{ij}}{ct_s} \;=\; \left(\hat{n}\cdot\frac{\vec{v}'}{c}\right)T_{ij}.$$

Since we assumed that the velocities of the sources are small ($|v'| \ll c$), we may neglect this term. Thus

$$\bar{h}_{ij} \;=\; -\frac{4G}{\kappa c^4 r}\int d^3x' T_{ij}\left(ct-r,\vec{x}'\right).$$

Within linearised gravity the Christoffel symbols are of order κ and therefore

$$0 \;=\; \nabla^{\mu}T_{\mu\nu} \;=\; \partial^{\mu}T_{\mu\nu} + \mathcal{O}(\kappa).$$

Thus

$$\begin{aligned}\partial^0 T_{00} + \partial^i T_{i0} &= 0,\\ \partial^0 T_{0j} + \partial^i T_{ij} &= 0.\end{aligned}$$

Combining the two equations we obtain

$$\left(\partial^0\right)^2 T_{00} \;=\; \partial^k\partial^l T_{kl}.$$

We multiply both sides by $x_i x_j$. Rearranging the right-hand side we obtain

$$\begin{aligned}\left(\partial^0\right)^2\left(T_{00}x_ix_j\right) &= \left(\partial^k\partial^l T_{kl}\right)x_ix_j \;=\; \partial^k\left[\left(\partial^l T_{kl}\right)x_ix_j\right] - \left(\partial^l T_{il}\right)x_j - \left(\partial^l T_{jl}\right)x_i\\ &= \partial^k\partial^l\left(T_{kl}x_ix_j\right) - \partial^k\left(T_{ki}x_j\right) - \partial^k\left(T_{kj}x_i\right) - \partial^l\left(T_{il}x_j\right) - \partial^l\left(T_{jl}x_i\right) + T_{ij} + T_{ji}\\ &= \partial^k\partial^l\left(T_{kl}x_ix_j\right) - 2\partial^k\left(T_{ki}x_j\right) - 2\partial^k\left(T_{kj}x_i\right) + 2T_{ij}.\end{aligned}$$

Thus

$$T_{ij} \;=\; \frac{1}{2}\left(\partial^0\right)^2\left(T_{00}x_ix_j\right) - \frac{1}{2}\partial^k\partial^l\left(T_{kl}x_ix_j\right) + \partial^k\left(T_{ki}x_j\right) + \partial^k\left(T_{kj}x_i\right).$$

The last three terms on the right-hand side are total derivatives with respect to the spatial coordinates. Plugging the expression for T_{ij} into our formula for $\bar{h}_{ij}$ we obtain

$$\begin{aligned}\bar{h}_{ij} &= -\frac{4G}{\kappa c^4 r}\int d^3x' T_{ij}\left(ct-r,\vec{x}'\right) \;=\; -\frac{2G}{\kappa c^4 r}\partial_0^2\int d^3x' T^{00}\left(ct-r,\vec{x}'\right)x'_ix'_j\\ &= -\frac{2G}{\kappa c^4 r}\partial_0^2\int d^3x' \rho\left(ct-r,\vec{x}'\right)x'_ix'_j.\end{aligned}$$

We obtain h_{ij}^{TT} with the help of the projection operator ${\Lambda_{ij}}^{kl}$

$$h_{ij}^{\mathrm{TT}} \;=\; {\Lambda_{ij}}^{kl}\bar{h}_{kl}.$$

The quadrupole moment is defined by

$$Q_{ij}\left(x^0\right) = \int d^3x' \rho\left(x^0, \vec{x}'\right)\left(3x'_i x'_j - r'^2 \delta_{ij}\right).$$

We finally obtain

$$h_{ij}^{\mathrm{TT}} = -\frac{2G}{3\kappa c^4 r}\Lambda_{ij}{}^{kl}\,\partial_0^2\,Q_{kl}\left(ct-r\right).$$

The term proportional to δ_{kl} in the quadrupole moment projects to zero. This formula is known as the **quadrupole formula**. Note the second time derivative: A static source cannot radiate gravitational waves. Note also that the source must possess at least a quadrupole moment. Monopole and dipole radiation is absent for gravitational waves.

Let us consider a simple example: A binary system of equal masses $m_1 = m_2 = m$ rotating on a circular orbit in the $x-y$-pane around the centre of mass.

$$\vec{x}_1(t) = \frac{r_b}{2}\begin{pmatrix}\cos\left(\omega_b t\right)\\ \sin\left(\omega_b t\right)\\ 0\end{pmatrix}, \qquad \vec{x}_2(t) = -\frac{r_b}{2}\begin{pmatrix}\cos\left(\omega_b t\right)\\ \sin\left(\omega_b t\right)\\ 0\end{pmatrix}.$$

Kepler's third law relates r_b and ω_b:

$$\omega_b^2 r_b^3 = 2Gm.$$

ρ is given by

$$\begin{aligned}&\rho\left(x\right) = mc^2\delta\left(z\right)\\ &\quad\times\left[\delta\left(x-\frac{r_b}{2}\cos\left(\omega_b t\right)\right)\delta\left(y-\frac{r_b}{2}\sin\left(\omega_b t\right)\right)+\delta\left(x+\frac{r_b}{2}\cos\left(\omega_b t\right)\right)\delta\left(y+\frac{r_b}{2}\sin\left(\omega_b t\right)\right)\right].\end{aligned}$$

The quadrupole moment is given by

$$Q_{ij}\left(ct\right) = \frac{1}{2}mc^2\frac{\left(2Gm\right)^{\frac{2}{3}}}{\omega_b^{\frac{4}{3}}}\begin{pmatrix}3\cos^2\left(\omega_b t\right)-1 & 3\cos\left(\omega_b t\right)\sin\left(\omega_b t\right) & 0\\ 3\cos\left(\omega_b t\right)\sin\left(\omega_b t\right) & 3\sin^2\left(\omega_b t\right)-1 & 0\\ 0 & 0 & -1\end{pmatrix}.$$

For the second time derivative we obtain

$$\partial_0^2\,Q_{ij} = -3m\left(2Gm\omega_b\right)^{\frac{2}{3}}\begin{pmatrix}\cos\left(2\omega_b t\right) & \sin\left(2\omega_b t\right) & 0\\ \sin\left(2\omega_b t\right) & -\cos\left(2\omega_b t\right) & 0\\ 0 & 0 & 0\end{pmatrix}.$$

Let us assume that the observer is placed along the z-direction at the distance z. Then

$$\kappa h_{ij}^{\mathrm{TT}} = \frac{\left(2Gm\right)^{\frac{5}{3}}\omega_b^{\frac{2}{3}}}{c^4 z}\begin{pmatrix}\cos\left(2\omega_b t-\phi_0\right) & \sin\left(2\omega_b t-\phi_0\right) & 0\\ \sin\left(2\omega_b t-\phi_0\right) & -\cos\left(2\omega_b t-\phi_0\right) & 0\\ 0 & 0 & 0\end{pmatrix},$$

where ϕ_0 is given by

$$\phi_0 = \frac{2\omega_b z}{c}.$$

This corresponds to a circular polarised gravitational wave. Note that the gravitational wave has angular frequency $2\omega_b$, where ω_b is the angular frequency of the rotating binary system.

For $m = m_\odot$, $\omega_b = 2\pi/(1\text{h})$ and $z = 1\text{kpc}$ one obtains

$$\frac{(2Gm)^{\frac{5}{3}}\,\omega_b^{\frac{2}{3}}}{c^4 z} \approx 10^{-21}.$$

We may repeat the exercise in the slightly more general situation for a binary system with unequal masses $m_1 \neq m_2$. It is convenient to introduce the total mass M and the reduced mass μ

$$M = m_1 + m_2, \qquad \mu = \frac{m_1 m_2}{m_1 + m_2}.$$

We may directly translate our previous formulae by noting that the energy density is proportional to the reduced mass μ, while the total mass enters Kepler's third law:

$$\omega_b^2 r_b^3 = GM.$$

Thus

$$\partial_0^2\, Q_{ij} = -6G^{\frac{2}{3}}\mu M^{\frac{2}{3}}\omega_b^{\frac{2}{3}} \begin{pmatrix} \cos(2\omega_b t) & \sin(2\omega_b t) & 0 \\ \sin(2\omega_b t) & -\cos(2\omega_b t) & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

and

$$\kappa h_{ij}^{\mathrm{TT}} = \frac{4G^{\frac{5}{3}}\mu M^{\frac{2}{3}}\omega_b^{\frac{2}{3}}}{c^4 z} \begin{pmatrix} \cos(2\omega_b t - \phi_0) & \sin(2\omega_b t - \phi_0) & 0 \\ \sin(2\omega_b t - \phi_0) & -\cos(2\omega_b t - \phi_0) & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

We see that these formulae only depend on the combination

$$M_c = \mu^{\frac{3}{5}} M^{\frac{2}{5}}.$$

M_c is called the **chirp mass**. In terms of the chirp mass we have

$$\partial_0^2\, Q_{ij} = -6G^{\frac{2}{3}} M_c^{\frac{5}{3}}\omega_b^{\frac{2}{3}} \begin{pmatrix} \cos(2\omega_b t) & \sin(2\omega_b t) & 0 \\ \sin(2\omega_b t) & -\cos(2\omega_b t) & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

and

$$\kappa h_{ij}^{\mathrm{TT}} = \frac{4c\omega_b^{\frac{2}{3}}}{z}\left(\frac{GM_c}{c^3}\right)^{\frac{5}{3}} \begin{pmatrix} \cos(2\omega_b t - \phi_0) & \sin(2\omega_b t - \phi_0) & 0 \\ \sin(2\omega_b t - \phi_0) & -\cos(2\omega_b t - \phi_0) & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

12.5 The energy-momentum tensor for gravitational waves

Up to now we treated gravitational waves in linearised gravity, expanded around flat Minkowski metric

$$g_{\mu\nu} = \eta_{\mu\nu} + \kappa h_{\mu\nu}.$$

We have seen that gravitational waves change the distance between two test masses and it is therefore clear that gravitational waves carry energy and momentum. We would like to determine the energy-momentum tensor associated to a gravitational wave. In order to this, we have to extend our formalism. There are two modifications required:

1. We have to set up a formalism, which allows an expansion around a curved background,
2. We have to expand to order κ^2.

Let us understand the first point: Within general relativity, any form of energy is a source of space-time curvature. A decomposition of the form $g_{\mu\nu} = \eta_{\mu\nu} + \kappa h_{\mu\nu}$ excludes from the very beginning the possibility that a gravitational wave deforms the background metric.

We therefore would like to write

$$g_{\mu\nu} = \bar{g}_{\mu\nu} + \kappa h_{\mu\nu},$$

where we think about $\bar{g}_{\mu\nu}$ as the background metric. However, in general such a decomposition is not unique. We have the problem of deciding which part belongs to $\bar{g}_{\mu\nu}$ and which part to $h_{\mu\nu}$. In order to have an un-ambiguous decomposition we need a hierarchy of scales: Let ω be the angular frequency of the gravitational wave and $\lambda\!\!\!^{-} = \lambda/(2\pi) = c/\omega$ the reduced wavelength. Denote by $\lambda\!\!\!^{-}_B$ the typical scale of the spatial variation of the background and by ω_B the typical angular frequency of the time variation of the background. Note that $\lambda\!\!\!^{-}_B$ and ω_B need not be related by $\lambda\!\!\!^{-}_B \omega_B = c$, they can be independent. We require

$$\lambda\!\!\!^{-} \ll \lambda\!\!\!^{-}_B, \qquad \omega \gg \omega_B.$$

Let's consider a classical analogy: Suppose that we are interested in water waves on the ocean. Take the reduced wavelength to be $\mathcal{O}(10^1\mathrm{m})$-$\mathcal{O}(10^2\mathrm{m})$. Typical wave velocities are $5-25\mathrm{ms}^{-1}$, giving $\omega \approx \mathcal{O}(1\mathrm{s}^{-1})$. The water waves propagate on a curved background: There is a spatial curvature due to the fact that the earth is a sphere, defining a length $\lambda\!\!\!^{-}_B = r_\oplus$ of the order of the earth's radius. In addition, there is a time variation due to tidal effects, defining an angular frequency $\omega_B \approx \mathcal{O}(1\mathrm{h}^{-1})$. For both cases we have a clear separation of scales.

Let us now assume a clear separation of scales: We assume $\lambda\!\!\!^{-} \ll \lambda\!\!\!^{-}_B$ and/or $\omega \gg \omega_B$. We may project to the background quantities by averaging: For $\lambda\!\!\!^{-} \ll \lambda\!\!\!^{-}_B$ we choose a l with $\lambda\!\!\!^{-} \ll l \ll \lambda\!\!\!^{-}_B$ and average over spatial volumes l^3. The short-wavelength modes of the gravitational waves will average out. For $\omega \gg \omega_B$ we choose a T with $1/\omega \ll T \ll 1/\omega_B$ and average over a time intervall of length T. Again, the high-frequency modes of the gravitational waves will average out. We denote the average by $\langle \dots \rangle$. A few examples for explicit averages (either over a time intervall or over a spatial volume or both) are

$$\begin{aligned} \left\langle \cos\left(\omega\left(t-\frac{r}{c}\right)\right)\right\rangle &= \left\langle \sin\left(\omega\left(t-\frac{r}{c}\right)\right)\right\rangle = 0, \\ \left\langle \cos^2\left(\omega\left(t-\frac{r}{c}\right)\right)\right\rangle &= \left\langle \sin^2\left(\omega\left(t-\frac{r}{c}\right)\right)\right\rangle = \frac{1}{2}. \end{aligned} \tag{1}$$

The average

$$\langle g_{\mu\nu} \rangle$$

gives the slowly-varying piece of the metric. We expand the full metric in κ:

$$g_{\mu\nu} = \bar{g}_{\mu\nu} + \kappa h_{\mu\nu} + \kappa^2 j_{\mu\nu} + \mathcal{O}\left(\kappa^3\right).$$

The lowest-order term $\bar{g}_{\mu\nu}$ is slowly-varying, while the first-order term $h_{\mu\nu}$ is rapidly-varying. The second-order term $j_{\mu\nu}$ has rapidly-varying and slowly-varying contributions. We split the second-order piece into a slowly-varying piece $\bar{j}_{\mu\nu}$ and a rapidly-varying piece $j_{\mu\nu}^{\mathrm{high}}$:

$$j_{\mu\nu} = \bar{j}_{\mu\nu} + j_{\mu\nu}^{\mathrm{high}}.$$

Separating the terms in the metric into slowly-varying / rapidly-varying contributions we have

$$g_{\mu\nu} = \left(\bar{g}_{\mu\nu} + \kappa^2 \bar{j}_{\mu\nu}\right) + \left(\kappa h_{\mu\nu} + \kappa^2 j_{\mu\nu}^{\mathrm{high}}\right) + \mathcal{O}\left(\kappa^3\right).$$

We would like to determine the contribution due to $\bar{j}_{\mu\nu}$. We expand the Einstein tensor in κ:

$$G_{\mu\nu} = G_{\mu\nu}^{(0)}\left(\bar{g}_{\rho\sigma}\right) + \kappa G_{\mu\nu}^{(1)}\left(\bar{g}_{\rho\sigma}, h_{\lambda\tau}\right) + \kappa^2 G_{\mu\nu}^{(1)}\left(\bar{g}_{\rho\sigma}, j_{\lambda\tau}\right) + \kappa^2 G_{\mu\nu}^{(2)}\left(\bar{g}_{\rho\sigma}, h_{\lambda\tau}\right) + \mathcal{O}\left(\kappa^3\right).$$

$G_{\mu\nu}^{(1)}$ and $G_{\mu\nu}^{(2)}$ can be obtained from a straightforward, but tedious second-order calculation.

Let us consider Einstein's equations in the vacuum. Einstein's equations hold order-by-order in κ:

$$\begin{aligned} G_{\mu\nu}^{(0)}\left(\bar{g}_{\rho\sigma}\right) &= 0, \\ G_{\mu\nu}^{(1)}\left(\bar{g}_{\rho\sigma}, h_{\lambda\tau}\right) &= 0, \\ G_{\mu\nu}^{(1)}\left(\bar{g}_{\rho\sigma}, j_{\lambda\tau}\right) + G_{\mu\nu}^{(2)}\left(\bar{g}_{\rho\sigma}, h_{\lambda\tau}\right) &= 0. \end{aligned}$$

We perform an average of the Einstein tensor:

$$0 = \left\langle G_{\mu\nu} \right\rangle = G_{\mu\nu}^{(0)}\left(\bar{g}_{\rho\sigma}\right) + \kappa^2 G_{\mu\nu}^{(1)}\left(\bar{g}_{\rho\sigma}, \bar{j}_{\lambda\tau}\right) + \kappa^2 \left\langle G_{\mu\nu}^{(2)}\left(\bar{g}_{\rho\sigma}, h_{\lambda\tau}\right)\right\rangle + \mathcal{O}\left(\kappa^3\right).$$

We re-write this equation as

$$G_{\mu\nu}^{(0)}\left(\bar{g}_{\rho\sigma}\right) + \kappa^2 G_{\mu\nu}^{(1)}\left(\bar{g}_{\rho\sigma}, \bar{j}_{\lambda\tau}\right) = -\kappa^2 \left\langle G_{\mu\nu}^{(2)}\left(\bar{g}_{\rho\sigma}, h_{\lambda\tau}\right)\right\rangle + \mathcal{O}\left(\kappa^3\right),$$

and define the effective energy-momentum tensor of a gravitational wave by

$$T_{\mu\nu}^{\mathrm{GW}} = -\frac{\kappa^2 c^4}{8\pi G} \left\langle G_{\mu\nu}^{(2)}\left(\bar{g}_{\rho\sigma}, h_{\lambda\tau}\right) \right\rangle.$$

It remain to calculate $T_{\mu\nu}^{\mathrm{GW}}$. One finds

$$T_{\mu\nu}^{\mathrm{GW}} = \frac{\kappa^2 c^4}{32\pi G} \left\langle \nabla_\mu \bar{h}_{\rho\sigma} \nabla_\nu \bar{h}^{\rho\sigma} - \frac{1}{2} \nabla_\mu \bar{h} \nabla_\nu \bar{h} - \nabla_\mu \bar{h}_{\nu\rho} \nabla_\sigma \bar{h}^{\rho\sigma} - \nabla_\nu \bar{h}_{\mu\rho} \nabla_\sigma \bar{h}^{\rho\sigma} \right\rangle,$$

where the covariant derivatives and the raising/lowering of indices are done with respect to $\bar{g}_{\mu\nu}$. (As we work to order $\mathcal{O}(\kappa^2)$ and this expression is already of order $\mathcal{O}(\kappa^2)$ anything else would be of higher order.) In the transverse traceless gauge this simplifies to

$$T_{\mu\nu}^{\mathrm{GW}} = \frac{\kappa^2 c^4}{32\pi G} \left\langle \nabla_\mu h_{\rho\sigma} \nabla_\nu h^{\rho\sigma} \right\rangle.$$

The energy density of a gravitational wave in the transverse traceless gauge is given by

$$\rho^{\mathrm{GW}} = T_{00}^{\mathrm{GW}} = \frac{\kappa^2 c^2}{32\pi G} \left\langle \dot{h}_{ij} \dot{h}^{ij} \right\rangle.$$

Let us now specialise to the case where

$$\bar{g}_{\mu\nu} = \eta_{\mu\nu}$$

and

$$\kappa h_{\mu\nu}^{\mathrm{TT}} = \kappa \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & h_+ & h_\times & 0 \\ 0 & h_\times & -h_+ & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \cos\left(\omega\left(t - \frac{z}{c}\right)\right).$$

We find

$$\rho^{\mathrm{GW}} = \frac{\kappa^2 c^2 \omega^2}{16\pi G} \left(h_+^2 + h_\times^2\right) \left\langle \cos^2\left(\omega\left(t - \frac{z}{c}\right)\right) \right\rangle = \frac{\kappa^2 c^2 \omega^2}{32\pi G} \left(h_+^2 + h_\times^2\right).$$

Let us now consider a binary system and let us work out the radiated energy per unit time (i.e. the radiation power). We start from

$$h_{ij}^{\mathrm{TT}} = -\frac{2G}{3\kappa c^6 r} \Lambda_{ij}{}^{kl} \, \ddot{Q}_{kl}\left(ct - r\right).$$

As in electrodynamics we define Poynting's vector (i.e. energy flux per unit time and unit area) by

$$S^k = c T^{0k} = \frac{\kappa^2 c^5}{32\pi G} \left\langle \partial^0 h_{ij} \partial^k h^{ij} \right\rangle.$$

It is convenient to use as spatial coordiantes (r, θ, ϕ). For a function $h_{ij} = (1/r) \cdot f_{ij}(ct-r)$ we have

$$\begin{aligned}
\frac{\partial}{\partial t} h_{ij} &= \frac{c}{r} f'_{ij}(ct-r), \\
\frac{\partial}{\partial r} h_{ij} &= -\frac{1}{r^2} f_{ij}(ct-r) - \frac{1}{r} f'_{ij}(ct-r) \;=\; -\frac{1}{c}\frac{\partial}{\partial t} h_{ij} + \mathcal{O}\left(\frac{1}{r^2}\right),
\end{aligned}$$

and hence

$$\partial_r h_{ij} \;=\; -\partial_0 h_{ij} + \mathcal{O}\left(\frac{1}{r^2}\right), \qquad \partial^r h_{ij} \;=\; \partial^0 h_{ij} + \mathcal{O}\left(\frac{1}{r^2}\right),$$

We obtain the radiated energy per unit time by integrating the energy flux per unit time and unit area over sphere with radius r:

$$\begin{aligned}
P &= r^2 \int d\Omega\, \vec{S} \cdot \hat{n} \;=\; r^2 \int d\Omega\, S^r \;=\; \frac{\kappa^2 c^5 r^2}{32\pi G} \int d\Omega\, \left\langle \partial^0 h_{ij} \partial^0 h^{ij} \right\rangle \\
&= \frac{G}{72\pi c^9} \left\langle \dddot{Q}_{kl}(ct-r)\; \dddot{Q}_{mn}(ct-r) \right\rangle \int d\Omega\, \Lambda_{ij}{}^{kl}\, \Lambda^{ijmn}
\end{aligned}$$

The angular integral gives

$$\int d\Omega\, \Lambda_{ij}{}^{kl}\, \Lambda^{ijmn} \;=\; \int d\Omega\, \Lambda^{klmn} \;=\; \frac{2\pi}{15}\left(11\delta^{km}\delta^{ln} - 4\delta^{kl}\delta^{mn} + \delta^{kn}\delta^{ml}\right).$$

Thus

$$P \;=\; \frac{G}{45c^9} \left\langle \dddot{Q}_{ij}(ct-r)\; \dddot{Q}^{ij}(ct-r) \right\rangle.$$

With

$$\dddot{Q}_{ij} \;=\; 12 c^2 G^{\frac{2}{3}} M_c^{\frac{5}{3}} \omega_b^{\frac{5}{3}} \begin{pmatrix} \sin(2\omega_b t - \phi_0) & -\cos(2\omega_b t - \phi_0) & 0 \\ -\cos(2\omega_b t - \phi_0) & -\sin(2\omega_b t - \phi_0) & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

we finally obtain

$$P \;=\; \frac{32}{5c^5} G^{\frac{7}{3}} M_c^{\frac{10}{3}} \omega_b^{\frac{10}{3}}.$$

12.6 The inspiral phase of a binary system

When we first derived the emission of gravitational waves from a binary system we assumed that the emission of gravitational waves has no impact on the binary system. In particular we assumed that the orbit is not changed. In reality this is not true. The gravitational waves carry away energy and momentum, causing the orbit of the binary system to shrink until coalescence. We may model the initial phase of the inspiral process with the tools we have up to now. The final phase of the inspiral process and the merger involve strong fields and cannot be described by perturbation theory. Here one resorts to numerical general relativity.

During the inspiral phase the following things happen:

- The total energy of the binary system decreases due to the emission of gravitational waves.
- This implies that r_b decreases and ω_b increases.
- If ω_b increases, the radiated power increases even more. This accelerates the process of energy-loss.

This will end with the coalescence.

We will model the beginning of the inspiral phase of the binary system by assuming that the orbit stays circular with a slowly decreasing radius ($|\dot{r}_S| \ll v$). The total energy of the binary system is ($M = m_1 + m_2$, $\mu = m_1 m_2/(m_1+m_2)$)

$$E_b \;=\; \langle T\rangle + \langle U\rangle \;=\; \frac{1}{2}\mu v^2 - \frac{G\mu M}{r_b} \;=\; -\frac{G\mu M}{2r_b},$$

where we used the virial theorem. Kepler's third law relates ω_b and r_b:

$$\omega_b^2 r_b^3 \;=\; GM,$$

and therefore

$$E_b \;=\; -\frac{1}{2}\mu G^{\frac{2}{3}} M^{\frac{2}{3}} \omega_b^{\frac{2}{3}} \;=\; -\frac{1}{2} G^{\frac{2}{3}} M_c^{\frac{5}{3}} \omega_b^{\frac{2}{3}}.$$

The loss of energy is given by the radiated power:

$$\frac{dE_b}{dt} \;=\; -P.$$

This leads to the equation

$$\dot{\omega}_b \;=\; \frac{96}{5}\left(\frac{GM_c}{c^3}\right)^{\frac{5}{3}} \omega_b^{\frac{11}{3}}.$$

The angular frequency ω of the gravitational wave is related to ω_b by $\omega = 2\omega_b$. We therefore have

$$\dot{\omega} \;=\; \frac{12\cdot 2^{\frac{1}{3}}}{5}\left(\frac{GM_c}{c^3}\right)^{\frac{5}{3}} \omega^{\frac{11}{3}}.$$

This equation allows us to determine the chirp mass of a binary system from the observation of the variation of the angular frequency of a gravitational wave.

We may integrate the differential equation and obtain

$$\omega(t) \;=\; \frac{5^{\frac{3}{8}}}{4}\left(\frac{GM_c}{c^3}\right)^{-\frac{5}{8}} \frac{1}{(t_c - t)^{\frac{3}{8}}},$$

where t_c denotes the time of coalescence. This expression is divergent at $t = t_c$, indicating that our perturbative treatment is not valid close to coalescence. The amplitude of the gravitational wave grows as

$$\frac{4c}{r}\left(\frac{GM_c}{c^3}\right)^{\frac{5}{3}} [\omega_b(t)]^{\frac{2}{3}} \;=\; 2^{\frac{4}{3}}\frac{c}{r}\left(\frac{GM_c}{c^3}\right)^{\frac{5}{3}} [\omega(t)]^{\frac{2}{3}}.$$

12.7 Post-Newtonian and post-Minkowskian expansions

Within perturbation theory we may systematically improve our predictions by including higher-order terms. Two formalisms are frequently used: The post-Newtonian and the post-Minkowskian expansion.

We start with the post-Newtonian expansion. In the dicussion of the production of gravitational waves we assumed weak gravitational fields and small velocities. If we consider again a binary system, we defined the Scharzschild radius as

$$r_s = \frac{2GM}{c^2},$$

where M denotes the total mass. The requirement of a weak gravitational field implies

$$\frac{r_s}{2r_b} \ll 1,$$

the requirement of small velocities implies

$$\left(\frac{v_b}{c}\right)^2 \ll 1.$$

The virial theorem relates the two small quantities:

$$\left(\frac{v_b}{c}\right)^2 = \frac{r_s}{2r_b}.$$

The **post-Newtonian expansion** is a simultaneous expansion in the two small quantities

$$\frac{v_b}{c} \quad \text{and} \quad \sqrt{\frac{r_s}{2r_b}},$$

where we treat v_b/c of the same order as $\sqrt{r_s/(2r_b)}$. This is an expansion in the weak gravitational field limit **and** the small velocity limit. When we derived the quadrupole formula for the emission of gravitational waves we basically worked in the lowest order of the post-Newtonian expansion.

For the **post-Minkowksian expansion** we only expand in the weak gravitational field limit. There are no restrictions on the velocities. The post-Minkowksian expansion is usually applied outside the source. We have

$$\sqrt{\frac{r_s}{2r_b}} = \frac{1}{c}\sqrt{\frac{GM}{r_b}}.$$

Outside the source we may treat M and r_b as fixed parameters and the post-Minkowskian expansion becomes an expansion in $\sqrt{G}$. When we discussed the propagation of gravitational waves and the detection of gravitational waves we basically worked in the lowest order of the post-Minkowskian expansion.

One usually employs the post-Newtonian expansion inside the source and the post-Minkowskian expansion outside the source. The reason is as follows: The post-Newtonian expansion is not valid far away from the source. To see this, we first note that for typical binary systems we have $r_b \ll \lambdabar$, i.e. a hierarchy

$$r_s \ll r_b \ll \lambdabar.$$

We call $r_b < r < \lambdabar$ the **near zone** and $\lambdabar < r$ the **far zone**. In the far zone, the metric perturbation is of the form

$$h_{\mu\nu} = \frac{1}{r} f_{\mu\nu}(ct - r).$$

Within the post-Newtonian expansion we reconstruct this function from its expansion for small retardations:

$$h_{\mu\nu} \approx \frac{1}{r}\left[f_{\mu\nu}(ct) - r\partial_0 f_{\mu\nu}(ct) + \frac{1}{2}r^2\partial_0^2 f_{\mu\nu}(ct) + \dots\right]$$

For

$$f_{\mu\nu}(ct - r) \sim \cos\left(\omega\left(t - \frac{r}{c}\right)\right) = \cos\left(\frac{ct - r}{\lambdabar}\right)$$

each derivative brings a factor $1/\lambdabar$. Within the post-Newtonian expansion we compute the retarded function as an expansion in $r/\lambdabar$. The expansion parameter is smaller one in the near zone, but not in the far zone. We do not expect the series expansion to converge in the far zone.

On the other hand, the post-Minkowksian expansion assumes only a weak gravitational field. If the gravitational field is weak inside the source (which we assume to be the case), then it is also weak outside the source and we may use the post-Minkowksian expansion down to $r > r_b$. (We do not use the post-Minkowksian expansion inside the source, the reason is simply that it is too complicated to keep the full velocity dependence.)

Thus the two expansions overlap in the near zone. In the near zone the predictions from the two expansions can be matched order by order in perturbation theory.

13 Perturbative quantum gravity

This chapter assumes a knowledge of quantum field theory.

13.1 Natural units

In quantum field theory it is common practice to use natural units

$$c = 1, \qquad \hbar = 1.$$

Furthermore it is common practice to rescale the fields and the sources. In the case of electrodynamics one rescales the fields and the sources as follows:

$$\begin{aligned}
\vec{E}^{\mathrm{nat}} &= \frac{1}{\sqrt{4\pi}} \vec{E}^{\mathrm{Gauss}}, & \rho^{\mathrm{nat}} &= \sqrt{4\pi}\rho^{\mathrm{Gauss}}, \\
\vec{B}^{\mathrm{nat}} &= \frac{1}{\sqrt{4\pi}} \vec{B}^{\mathrm{Gauss}}, & \vec{j}^{\mathrm{nat}} &= \sqrt{4\pi}\vec{j}^{\mathrm{Gauss}}.
\end{aligned}$$

Maxwell's equations in natural units (and with rescaled fields and sources) read

$$\begin{aligned}
\vec{\nabla} \cdot \vec{B} &= 0, & \vec{\nabla} \cdot \vec{E} &= \rho, \\
\vec{\nabla} \times \vec{E} + \partial_t \vec{B} &= 0, & \vec{\nabla} \times \vec{B} - \partial_t \vec{E} &= \vec{j}.
\end{aligned}$$

The Poisson equation in electrostatics reads

$$\Delta \Phi^{\mathrm{em}} = -\rho.$$

The Lagrange density of electrodynamics is given in natural units by

$$\mathscr{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu},$$

i.e. without an additional factor $1/(4\pi)$. The energy-momentum tensor of electrodynamics has in natural units likewise no explicit prefactor $1/(4\pi)$.

In this chapter we use natural units. Einstein's equations read in natural units

$$Ric_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R - \Lambda g_{\mu\nu} = 2G T_{\mu\nu}.$$

The action of general relativity reads in natural units

$$S_{EH} = -\frac{1}{4G} \int d^4x \, \sqrt{-g} \, (R + 2\Lambda).$$

We set

$$\kappa = \sqrt{8G},$$

and hence

$$S_{EH} = -\frac{2}{\kappa^2} \int d^4x \, \sqrt{-g} \, (R + 2\Lambda).$$

13.2 Low-energy effective theory of quantum gravity

In the derivation/motivation of Einstein's equations we considered the Newtonian limit. The Newtonian limit is defined by three conditions: (i) the gravitational field is weak, (ii) all velocities are small compared to the speed of ligth and (iii) the gravitational field is static. In this chapter we do not impose the last two conditions. We only assume that the gravitational field is weak. Thus we will treat the gravitational field as a small perturbation of the flat Minkowski metric.

Previously, we only considered classical physics, i.e. we looked at solutions of Einstein's equations (in the limit, where the gravitational field is weak). The path integral formalism allows us to go from classical physics to quantum physics: Instead of just considering the field configuration, which happens to be the solution of Einstein's equations, we now consider all possible field configurations and weight each field configuration by $\exp iS$. This gives us the low-energy effective theory of quantum gravity, which we may treat with perturbation theory. This gives us the correct quantum theory at low-energy. The effective theory breaks down at higher energies, where perturbations to the flat Minkowski metric no longer are small. The situation is similar to other effective theories like Fermi's four-fermion theory or chiral perturbation theory.

Within the low-energy effective theory we have a correspondence between gravitational waves and gravitons, in the same way as we have in quantum electrodynamics a correspondence between electromagnetic waves and photons. We may therefore discuss the scattering of gravitons. Let us stress that the experimental requirements for measuring the corresponding cross sections are far beyond the current experimental abilities. However, the discussion of graviton scattering amplitudes will reveal intriguing connections with Yang-Mills amplitudes.

We denote by

$$\eta_{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

the metric of flat Minkowski space. We write

$$g_{\mu\nu} = \eta_{\mu\nu} + \kappa h_{\mu\nu}$$

and treat $\kappa h_{\mu\nu}$ as a perturbation. We recall that we defined $\kappa = \sqrt{8G}$. The tensor $h_{\mu\nu}$ describes the graviton field. The metric $\eta_{\mu\nu}$ of flat Minkowski space is a solution of Einstein's equations without a cosmological constant:

$$Ric_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = 0.$$

We stress that $\eta_{\mu\nu}$ is not a solution of Einstein's equations with a non-zero cosmological constant. Our plan is to use perturbation theory around a solution of Einstein's euqations, therefore we restrict ourselves to $\Lambda = 0$. The Einstein-Hilbert action without a cosmological constant reads

$$S_{EH} = \int d^4x\, \mathscr{L}, \qquad \mathscr{L} = -\frac{2}{\kappa^2}\sqrt{-g}R.$$

We will treat $\kappa/4$ as a small coupling.

Let us now consider for $h_{\mu\nu}$ the effective (not necessarily renormalisable) quantum field theory described by the generating functional

$$Z\left[J^{\mu\nu}\right] = \int \mathcal{D}h_{\mu\nu} \, \exp\left[i\int d^4x \, \mathcal{L}_{\mathrm{EH}} + \mathcal{L}_{\mathrm{GF}} + \mathcal{L}_{\mathrm{FP}} + J^{\mu\nu}h_{\mu\nu}\right],$$

where $\mathcal{L}_{\mathrm{GF}}$ denotes the gauge-fixing term and $\mathcal{L}_{\mathrm{FP}}$ the corresponding Faddeev-Popov term. We will give an expression for the gauge-fixing term later on. The Faddeev-Popov term will only contribute to loop amplitudes. We will treat the quantum field theory defined by the equation above perturbatively. Our first goal is the expansion of the Lagrange density in powers of $h_{\mu\nu}$ (or equivalently in powers of κ). Let us introduce the following abbreviations:

$$\begin{aligned}
(\eta h\eta)^{\mu\nu} &= \eta^{\mu\mu_1}h_{\mu_1\mu_2}\eta^{\mu_2\nu}, \\
(\eta h\eta h\eta)^{\mu\nu} &= \eta^{\mu\mu_1}h_{\mu_1\mu_2}\eta^{\mu_2\mu_3}h_{\mu_3\mu_4}\eta^{\mu_4\nu}, \\
(\eta h\eta h\eta h\eta)^{\mu\nu} &= \eta^{\mu\mu_1}h_{\mu_1\mu_2}\eta^{\mu_2\mu_3}h_{\mu_3\mu_4}\eta^{\mu_4\mu_5}h_{\mu_5\mu_6}\eta^{\mu_6\nu}.
\end{aligned}$$

With the help of these abbreviations we may express the inverse metric tensor $g^{\mu\nu}$ through $h_{\mu\nu}$:

$$g^{\mu\nu} = \eta^{\mu\nu} - \kappa(\eta h\eta)^{\mu\nu} + \kappa^2(\eta h\eta h\eta)^{\mu\nu} - \kappa^3(\eta h\eta h\eta h\eta)^{\mu\nu} + \mathcal{O}\left(\kappa^4\right).$$

The inverse metric tensor is an infinite power series in κ. Let us now turn to the determinant $g = \det(g_{\mu\nu})$. Also here we introduce a few abbreviations:

$$\begin{aligned}
(\eta h) &= \eta^{\mu_1\mu_2}h_{\mu_2\mu_1}, \\
(\eta h\eta h) &= \eta^{\mu_1\mu_2}h_{\mu_2\mu_3}\eta^{\mu_3\mu_4}h_{\mu_4\mu_1}, \\
(\eta h\eta h\eta h) &= \eta^{\mu_1\mu_2}h_{\mu_2\mu_3}\eta^{\mu_3\mu_4}h_{\mu_4\mu_5}\eta^{\mu_5\mu_6}h_{\mu_6\mu_1}, \\
(\eta h\eta h\eta h\eta h) &= \eta^{\mu_1\mu_2}h_{\mu_2\mu_3}\eta^{\mu_3\mu_4}h_{\mu_4\mu_5}\eta^{\mu_5\mu_6}h_{\mu_6\mu_7}\eta^{\mu_7\mu_8}h_{\mu_8\mu_1}.
\end{aligned}$$

We then find for the determinant:

$$\begin{aligned}
&-\det\left(g_{\mu\nu}\right) = \\
&\quad 1 + \kappa(\eta h) + \kappa^2\left[\frac{1}{2}(\eta h)^2 - \frac{1}{2}(\eta h\eta h)\right] + \kappa^3\left[\frac{1}{6}(\eta h)^3 - \frac{1}{2}(\eta h\eta h)(\eta h) + \frac{1}{3}(\eta h\eta h\eta h)\right] \\
&\quad +\kappa^4\left[\frac{1}{24}(\eta h)^4 - \frac{1}{4}(\eta h\eta h)(\eta h)^2 + \frac{1}{8}(\eta h\eta h)^2 + \frac{1}{3}(\eta h\eta h\eta h)(\eta h) - \frac{1}{4}(\eta h\eta h\eta h\eta h)\right].
\end{aligned}$$

Note that this expression is a polynomial in κ and terminates with the κ^4-term. However, by taking the square root of this expression we again obtain an infinite power series in κ:

$$\begin{aligned}
\sqrt{-g} &= 1 + \frac{\kappa}{2}(\eta h) + \frac{\kappa^2}{8}\left[(\eta h)^2 - 2(\eta h\eta h)\right] + \frac{\kappa^3}{48}\left[(\eta h)^3 - 6(\eta h\eta h)(\eta h) + 8(\eta h\eta h\eta h)\right] \\
&\quad +\mathcal{O}\left(\kappa^4\right).
\end{aligned}$$

In order to find the expression for the scalar curvature R let us first consider the Christoffel symbols

$$\Gamma_{\kappa\mu\nu} = \frac{1}{2}\left(\partial_\mu g_{\nu\kappa} + \partial_\nu g_{\mu\kappa} - \partial_\kappa g_{\mu\nu}\right) = \frac{\kappa}{2}\left(\partial_\mu h_{\nu\kappa} + \partial_\nu h_{\mu\kappa} - \partial_\kappa h_{\mu\nu}\right).$$

Here we used $\partial_\alpha \eta_{\beta\gamma} = 0$. The Riemann curvature tensor is then given by

$$R_{\kappa\lambda\mu\nu} = \frac{\kappa}{2}\left(\partial_\lambda\partial_\mu h_{\kappa\nu} - \partial_\kappa\partial_\mu h_{\lambda\nu} + \partial_\kappa\partial_\nu h_{\lambda\mu} - \partial_\lambda\partial_\nu h_{\kappa\mu}\right) + g^{\xi\eta}\left(\Gamma_{\xi\kappa\nu}\Gamma_{\eta\lambda\mu} - \Gamma_{\xi\kappa\mu}\Gamma_{\eta\lambda\nu}\right).$$

The first term is linear in $h_{\mu\nu}$, while the second term is at least quadratic in $h_{\mu\nu}$. For the scalar curvature we have then

$$R = g^{\kappa\mu} g^{\lambda\nu} R_{\kappa\lambda\mu\nu}.$$

Since both $g^{\mu\nu}$ and $\sqrt{-g}$ are infinite power series in κ we obtain for the Lagrange density an infinite power series in κ as well. We write

$$\mathscr{L}_{\text{EH}} + \mathscr{L}_{\text{GF}} = \sum_{j=1}^{\infty} \mathscr{L}^{(j)},$$

where the term $\mathscr{L}^{(j)}$ contains the field $h_{\mu\nu}$ exactly j times. In this way we obtain a theory with an infinite tower of vertices, ordered by the number of the fields. The term $\mathscr{L}^{(1)}$ is given by

$$\mathscr{L}^{(1)} = -\frac{2}{\kappa}\eta^{\kappa\mu}\eta^{\lambda\nu}\partial_\lambda\left(\partial_\mu h_{\kappa\nu} - \partial_\nu h_{\kappa\mu}\right).$$

This term is a total derivative and vanishes in the action after partial integration:

$$-\frac{2}{\kappa}\eta^{\kappa\mu}\eta^{\lambda\nu}\int d^4x\, \partial_\lambda\left(\partial_\mu h_{\kappa\nu} - \partial_\nu h_{\kappa\mu}\right) = 0.$$

We may therefore ignore this term and start the expansion of the Lagrange density in powers of κ with the term quadratic in $h_{\mu\nu}$.

Let us add the following remark: If we would have expanded naively the Einstein-Hilbert action with a cosmological constant $\Lambda \neq 0$ around the flat Minkowski metric $\eta_{\mu\nu}$, we would have picked up an additional term

$$-\frac{2\Lambda}{\kappa}\eta^{\mu\nu}h_{\mu\nu}$$

contributing to $\mathscr{L}^{(1)}$, coming from the expansion of $\sqrt{-g}$. This additional term is not a total derivative and does not vanish. Terms of this type are called tadpoles and indicate that we expanded around the wrong background field.

Let us now return to the case $\Lambda = 0$. We consider the term $\mathscr{L}^{(2)}$, bilinear in $h_{\mu\nu}$. The gauge-fixing term $\mathscr{L}_{\mathrm{GF}}$ gives a contribution to $\mathscr{L}^{(2)}$. A popular gauge choice for gravity is **de Donder gauge**. This gauge is defined by

$$\mathscr{L}_{\mathrm{GF}} = \frac{1}{\kappa^2} C_\mu \eta^{\mu\nu} C_\nu,$$

where C_μ is given by

$$C_\mu = \eta^{\alpha\beta}\Gamma_{\mu\alpha\beta} = \frac{\kappa}{2}\eta^{\alpha\beta}\left(\partial_\alpha h_{\beta\mu} + \partial_\beta h_{\alpha\mu} - \partial_\mu h_{\alpha\beta}\right) = \kappa\eta^{\alpha\beta}\left(\partial_\alpha h_{\beta\mu} - \frac{1}{2}\partial_\mu h_{\alpha\beta}\right).$$

In this gauge one finds

$$\mathscr{L}^{(2)} = \frac{1}{2} h_{\mu_1\mu_2}\left(\frac{1}{2}\eta^{\mu_1\mu_2}\eta^{\nu_1\nu_2} - \frac{1}{2}\eta^{\mu_1\nu_1}\eta^{\mu_2\nu_2} - \frac{1}{2}\eta^{\mu_1\nu_2}\eta^{\mu_2\nu_1}\right)\Box h_{\nu_1\nu_2}.$$

Here, we symmetrised the expression in the bracket in (μ_1,μ_2) and (ν_1,ν_2). We are free to do this, since $h_{\mu\nu}$ is symmetric under an exchange of μ and ν. Let us first consider the tensor structure (in D space-time dimensions). For

$$\begin{aligned} M^{\mu_1\mu_2\nu_1\nu_2} &= \frac{1}{2}\eta^{\mu_1\nu_1}\eta^{\mu_2\nu_2} + \frac{1}{2}\eta^{\mu_1\nu_2}\eta^{\mu_2\nu_1} - \frac{1}{2}\eta^{\mu_1\mu_2}\eta^{\nu_1\nu_2}, \\ N_{\mu_1\mu_2\nu_1\nu_2} &= \frac{1}{2}\left(\eta_{\mu_1\nu_1}\eta_{\mu_2\nu_2} + \eta_{\mu_1\nu_2}\eta_{\mu_2\nu_1} - \frac{2}{D-2}\eta_{\mu_1\mu_2}\eta_{\nu_1\nu_2}\right) \end{aligned}$$

we have

$$M^{\mu_1\mu_2\rho_1\rho_2} N_{\rho_1\rho_2\nu_1\nu_2} = \frac{1}{2}\left(\delta^{\mu_1}_{\nu_1}\delta^{\mu_2}_{\nu_2} + \delta^{\mu_1}_{\nu_2}\delta^{\mu_2}_{\nu_1}\right).$$

The propagator of the graviton is therefore given by

$$\frac{1}{2}\left(\eta_{\mu_1\nu_1}\eta_{\mu_2\nu_2} + \eta_{\mu_1\nu_2}\eta_{\mu_2\nu_1} - \frac{2}{D-2}\eta_{\mu_1\mu_2}\eta_{\nu_1\nu_2}\right)\frac{i}{p^2}.$$

Let us now turn to the three-graviton vertex. The three-graviton vertex is determined by $\mathscr{L}^{(3)}$. After a longer calculation and by using integration-by-parts one finds

$$\begin{aligned} \mathscr{L}^{(3)} = \kappa\Big[& -\frac{1}{4}\eta^{\mu_1\nu_1}\eta^{\mu_2\nu_2}\eta^{\mu_3\nu_3}\eta^{\rho_2\rho_3} + \frac{1}{4}\eta^{\mu_1\nu_1}\eta^{\mu_2\nu_3}\eta^{\mu_3\nu_2}\eta^{\rho_2\rho_3} + \eta^{\mu_1\nu_2}\eta^{\mu_2\nu_1}\eta^{\mu_3\nu_3}\eta^{\rho_2\rho_3} \\ & - \eta^{\mu_1\nu_2}\eta^{\mu_2\nu_3}\eta^{\mu_3\nu_1}\eta^{\rho_2\rho_3} + \frac{1}{2}\eta^{\mu_1\rho_2}\eta^{\rho_3\nu_1}\eta^{\mu_2\nu_2}\eta^{\mu_3\nu_3} - \frac{1}{2}\eta^{\mu_1\rho_2}\eta^{\rho_3\nu_1}\eta^{\mu_2\nu_3}\eta^{\mu_3\nu_2} \\ & + 2\eta^{\mu_1\rho_2}\eta^{\rho_3\nu_2}\eta^{\mu_2\nu_3}\eta^{\mu_3\nu_1} - \eta^{\mu_1\rho_2}\eta^{\rho_3\nu_2}\eta^{\mu_2\nu_1}\eta^{\mu_3\nu_3} - \frac{1}{2}\eta^{\mu_3\rho_2}\eta^{\rho_3\nu_2}\eta^{\mu_1\nu_1}\eta^{\mu_2\nu_3} \\ & + \eta^{\mu_3\rho_2}\eta^{\rho_3\nu_2}\eta^{\mu_1\nu_3}\eta^{\mu_2\nu_1} - \eta^{\mu_1\rho_2}\eta^{\rho_3\nu_3}\eta^{\mu_2\nu_2}\eta^{\mu_3\nu_1} - \eta^{\mu_3\rho_2}\eta^{\rho_3\nu_3}\eta^{\mu_1\nu_2}\eta^{\mu_2\nu_1} \\ & + \frac{1}{2}\eta^{\mu_3\rho_2}\eta^{\rho_3\nu_3}\eta^{\mu_1\nu_1}\eta^{\mu_2\nu_2}\Big] h_{\mu_1\nu_1}\left(\partial_{\rho_2} h_{\mu_2\nu_2}\right)\left(\partial_{\rho_3} h_{\mu_3\nu_3}\right). \end{aligned}$$

Let us write $\mathcal{L}^{(3)}$ as

$$\mathcal{L}^{(3)} = O^{\mu_1\mu_2\mu_3\nu_1\nu_2\nu_3}\left(\partial_1,\partial_2,\partial_3\right) h_{\mu_1\nu_1} h_{\mu_2\nu_2} h_{\mu_3\nu_3},$$

where $O^{\mu_1\mu_2\mu_3\nu_1\nu_2\nu_3}\left(\partial_1,\partial_2,\partial_3\right)$ is defined by comparison with the previous equation (2). The notation ∂_j denotes a derivative acting on the field $h_{\mu_j\nu_j}$. The Feynman rule for the three-graviton vertex is then

$$V^{\mu_1\mu_2\mu_3\nu_1\nu_2\nu_3}\left(p_1,p_2,p_3\right) = i \sum_{\sigma\in S_3} O^{\mu_{\sigma(1)}\mu_{\sigma(2)}\mu_{\sigma(3)}\nu_{\sigma(1)}\nu_{\sigma(2)}\nu_{\sigma(3)}}\left(ip_{\sigma(1)},ip_{\sigma(2)},ip_{\sigma(3)}\right).$$

The explicit expression for $V^{\mu_1\mu_2\mu_3\nu_1\nu_2\nu_3}$ is rather long and not given here. However, one interesting property should be mentioned: The three-graviton vertex can be written as

$$V^{\mu_1\mu_2\mu_3\nu_1\nu_2\nu_3}\left(p_1,p_2,p_3\right) = i\,\frac{\kappa}{4}\, V^{\mu_1\mu_2\mu_3}\left(p_1,p_2,p_3\right) V^{\nu_1\nu_2\nu_3}\left(p_1,p_2,p_3\right) + ...,$$

where the dots denote terms, which vanish in the on-shell limit. The expression $V^{\mu_1\mu_2\mu_3}\left(p_1,p_2,p_3\right)$ is the Feynman rule for the colour-stripped cyclic-order three-gluon vertex, given by

$$V^{\mu_1\mu_2\mu_3}\left(p_1,p_2,p_3\right) = i\left[g^{\mu_1\mu_2}\left(p_1^{\mu_3}-p_2^{\mu_3}\right)+g^{\mu_2\mu_3}\left(p_2^{\mu_1}-p_3^{\mu_1}\right)+g^{\mu_3\mu_1}\left(p_3^{\mu_2}-p_1^{\mu_2}\right)\right].$$

We see that the three-graviton vertex in the on-shell limit is given (up to a prefactor involving the coupling) as the square of the cyclic-ordered three gluon vertex. This relates gravity with non-abelian gauge theories and is known as the **double-copy** property.

In principle it is possible to derive from the Lagrange density systematically the additional Feynman rules for vertices with four, five, ..., n gravitons. In addition we need a rule for the external graviton states. This rule is rather simple. A graviton is a spin 2 particle with two polarisation states, corresponding to the helicities $h = +2$ and $h = -2$. We label these states by $++$ and $--$. We may describe the polarisation tensor of an external graviton by a product of two polarisation vectors for gauge bosons:

$$\varepsilon^{++}_{\mu\nu}(p) = \varepsilon^{+}_{\mu}(p)\,\varepsilon^{+}_{\nu}(p), \qquad \varepsilon^{--}_{\mu\nu}(p) = \varepsilon^{-}_{\mu}(p)\,\varepsilon^{-}_{\nu}(p).$$

For the calculation of the scattering amplitude with n gravitons we will need all vertices with up to n gravitons. The scattering amplitude may then be computed through Feynman diagrams. However, this approach is rather tedious. More efficient methods are based on the "double-copy"-property or on-shell recursion formulae.

13.3 Interaction of gravitons with matter

We will model matter by a massive (complex) scalar field. The relevant Lagrangian for the coupling of a complex scalar field to gravity is given by

$$\mathcal{L}_{\text{scalar}} = \sqrt{-g}\left[\left(\partial_\mu\phi^*\right)\left(\partial_\nu\phi\right) g^{\mu\nu} - m^2\phi^*\phi\right].$$

As previously, we expand this Lagrange density in a series in κ:

$$\mathcal{L}_{\text{scalar}} = \sum_{i=0}^{\infty} \mathcal{L}_{\text{scalar}}^{(i)}.$$

The zeroth-order term $\mathcal{L}_{\text{scalar}}^{(0)}$ reads

$$\mathcal{L}_{\text{scalar}}^{(0)} = \left(\partial_\mu \phi^*\right)\left(\partial_\nu \phi\right)\eta^{\mu\nu} - m^2 \phi^* \phi.$$

This term gives the propagator of the scalar field:

$$\frac{i}{p^2 - m^2}.$$

The term $\mathcal{L}_{\text{scalar}}^{(1)}$ reads

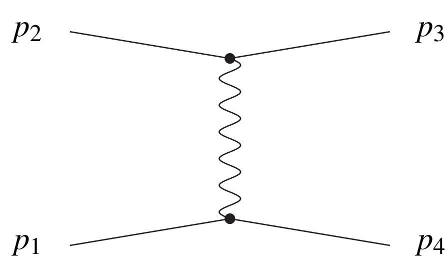
$$\mathcal{L}_{\text{scalar}}^{(1)} = \frac{\kappa}{4}\left[2\left(\eta^{\mu_1\mu_2}\eta^{\mu_3\mu_4} - \eta^{\mu_1\mu_3}\eta^{\mu_2\mu_4} - \eta^{\mu_1\mu_4}\eta^{\mu_2\mu_3}\right) h_{\mu_1\mu_2}\left(\partial_{\mu_3}\phi^*\right)\left(\partial_{\mu_4}\phi\right) - 2m^2\eta^{\mu_1\mu_2} h_{\mu_1\mu_2}\phi^*\phi\right].$$

From this term we derive the Feynman rule for the scalar-scalar-graviton vertex:

$$i\frac{\kappa}{4}\left[2p_1^{\mu_1}p_2^{\mu_2} + 2p_2^{\mu_1}p_1^{\mu_2} - \left(2p_1 \cdot p_2 + 2m^2\right)\eta^{\mu_1\mu_2}\right],$$

where p_1 denotes the momentum of the outgoing ϕ^*-particle and p_2 denotes the momentum of the outgoing ϕ-particle.

We may now calculate the scattering amplitude for the scattering of two scalar particles with masses m and m' through the exchange of a graviton. Theres is only one Feynman diagram:



We obtain for the scattering amplitude

$$\begin{aligned}\mathcal{M} = & \left(\frac{\kappa}{4}\right) i\left[2p_2^{\mu_1}p_3^{\mu_2} + 2p_3^{\mu_1}p_2^{\mu_2} - \left(2p_2p_3 + 2m'^2\right)\eta^{\mu_1\mu_2}\right] \\ & \times \frac{1}{2}\left[\eta_{\mu_1\nu_1}\eta_{\mu_2\nu_2} + \eta_{\mu_1\nu_2}\eta_{\mu_2\nu_1} - \eta_{\mu_1\mu_2}\eta_{\nu_1\nu_2}\right]\frac{i}{\left(p_2+p_3\right)^2} \\ & \times \left(\frac{\kappa}{4}\right) i\left[2p_1^{\nu_1}p_4^{\nu_2} + 2p_4^{\nu_1}p_1^{\nu_2} - \left(2p_1p_4 + 2m^2\right)\eta^{\nu_1\nu_2}\right],\end{aligned}$$

where the first line contains the Feynman rule for the upper scalar-scalar-graviton vertex, the second line contains the Feynman rule for the graviton propagator and the third line contains the Feynman rule for the lower scalar-scalar-graviton vertex. The contraction of the indices leads to

$$\mathcal{M} = -i\left(\frac{\kappa}{4}\right)^2 \frac{4}{t}\left[(s+u)\left(m^2+m'^2\right) - su - m^4 - m'^4 - 4m^2m'^2\right].$$

Here we introduced the Mandelstam variables

$$s = (p_1+p_2)^2 = (p_3+p_4)^2, \quad t = (p_2+p_3)^2 = (p_1+p_4)^2, \quad u = (p_1+p_3)^2 = (p_2+p_4)^2.$$

Let us now consider the scattering process $\phi_1\phi_2 \to \phi_3\phi_4$ in the non-relativistic limit. In this limit the spatial components of the four-vectors are small against the energy components. If we only keep the leading term of each component we have

$$p_1^\mu = (-m, -\vec{p}_1), \quad p_2^\mu = (-m', -\vec{p}_2), \quad p_3^\mu = (m', \vec{p}_3), \quad p_4^\mu = (m, \vec{p}_4).$$

The minus sign in p_1 and p_2 is related to the fact that within our convention we consider all momenta as outgoing. For the Mandelstam variables s and u we obtain

$$s = \left(m+m'\right)^2, \quad u = \left(m-m'\right)^2.$$

For the Mandelstam variable t we obtain

$$t = -|\vec{p}_3 - \vec{p}_2|^2 = -|\vec{p}_4 - \vec{p}_1|^2 = -|\vec{q}|^2.$$

In the non-relativistic limit the Mandelstam variable t is small against all other variables s, u, m^2 und m'^2. Thus, we may neglect t in the numerator of the scattering amplitude. In the non-relativistic limit the scattering amplitude simplifies to

$$\mathcal{M} = i\left(\frac{\kappa}{4}\right)^2 \frac{8m^2m'^2}{|\vec{q}|^2} = 4i\frac{Gm^2m'^2}{|\vec{q}|^2}.$$

Let us compare this scattering amplitude to the scattering amplitude for the scattering of two electrically charged fermions with charges Q and Q' and masses m and m'. Within quantum electrodynamics we obtain in the non-relativistic limit

$$\mathcal{A} = -4i\frac{QQ'mm'}{|\vec{q}|^2}.$$

Let us first consider the signs. From electrodynamics we know that equal-sign charges ($QQ' > 0$) repel each other, while opposite-sign charges ($QQ' < 0$) attract each other. From the sign of $\mathcal{M}$ we conclude that gravitation is always an attractive force.

The two scattering amplitudes agree up to prefactors. The kinematic dependence on the momenta is given in both cases by the factor $1/|\vec{q}|^2$ and corresponds in the classicial limit to an $1/r$-potential.

13.4 The relation between graviton amplitudes and Yang-Mills amplitudes

We finish this lecture with a remarkable relation between scattering amplitudes in three – at first-sight – different theories. We consider (i) gravity specified by

$$\mathcal{L}_{EH} = -\frac{2}{\kappa^2}\sqrt{-g}R,$$

(ii) Yang-Mills theory specified by the Lagrangian

$$\mathcal{L}_{YM} = -\frac{1}{4}F^a_{\mu\nu}F^{a\,\mu\nu},$$

and (iii) a bi-adjoint scalar theory specified by the Lagrangian

$$\mathcal{L}_{\mathrm{bi-adjoint\ scalar}} = \frac{1}{2}\left(\partial_\mu\phi^{ab}\right)\left(\partial^\mu\phi^{ab}\right) - \frac{\lambda}{3!}f^{a_1a_2a_3}\tilde{f}^{b_1b_2b_3}\phi^{a_1b_1}\phi^{a_2b_2}\phi^{a_3b_3}.$$

Let us first comment on the last two theories: We start with Yang-Mill theory. This is a gauge theory. Gauge theories describe the strong, weak and electromagnetic interactions. We denote by G the gauge group, this is a Lie group. We consider a non-Abelian gauge group (an example could be $SU(3)$, which is relevant for the strong interactions). We denote $\mathfrak{g}$ its Lie algebra and T^a the generators of the Lie algebra where the index a takes values from 1 to $\dim G$. We use the conventions

$$\left[T^a, T^b\right] = i f^{abc} T^c, \qquad \mathrm{Tr}\left(T^a T^b\right) = \frac{1}{2}\delta^{ab}.$$

We denote by $A^a_\mu(x)$ the gauge field. The field describes a massless spin-1 boson. The field strength is given by

$$F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g f^{abc} A^b_\mu A^c_\nu.$$

The coupling of Yang-Mills theory is denoted by g. The Lagrange density is invariant under **local gauge transformations**

$$T^a A^a_\mu(x) \rightarrow U(x)\left(T^a A^a_\mu(x) + \frac{i}{g}\partial_\mu\right)U^\dagger(x),$$

with

$$U(x) = \exp\left(-iT^a\theta_a(x)\right).$$

Let us now consider scattering amplitudes of n gauge bosons to lowest in perturbation theory. These amplitudes depend on a set of n four-vectors $p = (p_1, p_2, \dots, p_n)$, describing the momenta of the n gauge bosons and a set of n polarisation vectors $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)$, describing the spins/polarisations of the n gauge bosons. A gauge boson is a spin 1 particle and has two spin states, either the projection of the spin along the momentum is $+1$ (positive helicity) or -1 (negative helicity). We denote the corresponding polarisation vectors by ε^+_μ and ε^-_μ. We denote the tree amplitude by $\mathcal{A}_n^{(0)}(p,\varepsilon)$. We may write the amplitude in a form, where we group terms with the same group-theoretical factors together:

$$\mathcal{A}_n^{(0)}(p,\varepsilon) = g^{n-2} \sum_{\sigma \in S_n/Z_n} 2\,\mathrm{Tr}\left(T^{a_{\sigma(1)}} \dots T^{a_{\sigma(n)}}\right) \; A_n^{(0)}(\sigma,p,\varepsilon).$$

The expression on the right-hand side is called the **colour-decomposition** of the Yang-Mills amplitude. The quantities $A_n^{(0)}(\sigma,p,\varepsilon)$ accompanying the colour factor $2\,\mathrm{Tr}(T^{a_{\sigma(1)}} \dots T^{a_{\sigma(n)}})$ are called **partial amplitudes**. Partial amplitudes are gauge-invariant. Closely related are **primitive amplitudes**, which for tree-level Yang-Mills amplitudes are calculated from planar diagrams with a fixed cyclic ordering of the external legs and cyclic-ordered Feynman rules. Primitive amplitudes are gauge invariant as well. For tree-level Yang-Mills amplitudes the notions of partial amplitudes and primitive amplitudes coincide. Primitive amplitudes depend on p, ε and a permutation $\sigma \in S_n$. Let us now keep p and ε fixed and view $A_n^{(0)}(\sigma,p,\varepsilon)$ as a function of σ. For simplicity we suppress the dependence on p and ε and write

$$A_n^{(0)}(\sigma_1,\dots,\sigma_n) = A_n^{(0)}(\sigma,p,\varepsilon).$$

An obvious question related to the colour decomposition is: How many independent primitive amplitudes are there for n external particles? For a fixed set of external momenta and a fixed set of polarisations the primitive amplitudes are distinguished by the permutation specifying the order of the external particles. For n external particles there are $n!$ permutations and therefore $n!$ different orders. However, there are relations among primitive amplitudes with different external order. The first set of relations is rather trivial and given by **cyclic invariance**:

$$A_n^{(0)}(1,2,...,n) = A_n^{(0)}(2,...,n,1)$$

Cyclic invariance is the statement that only the external cyclic order matters, not the point, where we start to read off the order. Cyclic invariance reduces the number of independent primitive amplitudes to $(n-1)!$.

The first non-trivial relations are the **Kleiss-Kuijf relations**. Let

$$\vec{\alpha} = \left(\alpha_1,\alpha_2,...,\alpha_j\right), \qquad \vec{\beta} = \left(\beta_1,\beta_2,...,\beta_{n-2-j}\right)$$

be two ordered sequences of numbers, such that

$$\{1\} \cup \{\alpha_1,...,\alpha_j\} \cup \{\beta_1,...,\beta_{n-2-j}\} \cup \{n\} = \{1,...,n\}.$$

We further set $\vec{\beta}^T = (\beta_{n-2-j},...,\beta_2,\beta_1)$. The Kleiss-Kuijf relations read

$$A_n^{(0)}(1,\alpha_1,...,\alpha_j,n,\beta_1,...,\beta_{n-2-j}) = (-1)^{n-2-j} \sum_{\sigma \in \vec{\alpha} \sqcup\!\sqcup \vec{\beta}^T} A_n^{(0)}(1,\sigma_1,...,\sigma_{n-2},n).$$

Here, $\vec{\alpha} \sqcup\!\sqcup \vec{\beta}^T$ denotes the set of all shuffles of $\vec{\alpha}$ with $\vec{\beta}^T$, i.e. the set of all permutations of the elements of $\vec{\alpha}$ and $\vec{\beta}^T$, which preserve the relative order of the elements of $\vec{\alpha}$ and of the elements of $\vec{\beta}^T$. The Kleiss-Kuijf relations reduce the number of independent primitive amplitudes to $(n-2)!$.

Apart from cyclic invariance and the Kleiss-Kuijf relations there are in addition the **Bern-Carrasco-Johansson relations** (BCJ relations). The fundamental BCJ relations read

$$\sum_{i=2}^{n-1}\left(\sum_{j=i+1}^{n} 2p_2p_j\right) A_n^{(0)}(1,3,...,i,2,i+1,...,n-1,n) = 0.$$

Cyclic invariance allows us to fix one external particle at a specified position, say position 1. The Kleiss-Kuijf relations allow us to fix a second external particle at another specified position, say position n. The BCJ relations allow us to fix a third external particle at a third specified position, say position 2. The BCJ relations reduce the number of independent primitive amplitudes to $(n-3)!$. The full set of relations among primitive tree amplitudes in pure Yang-Mills theory is given by cyclic invariance, Kleiss-Kuijf relations, and the fundamental BCJ relations. Therefore a basis of independent primitive amplitudes consists of $(n-3)!$ elements.

Let us now turn to the bi-adjoint scalar theory. This theory consists of a scalar field ϕ^{ab} in adjoint representation of two Lie groups G and $\tilde{G}$. We will denote indices referring to G by a, indices referring to $\tilde{G}$ by b. Amplitudes in this theory have a double colour decomposition, similar to the (single) colour decomposition of gauge amplitudes:

$$\mathcal{m}_n^{(0)}(p) = \lambda^{n-2} \sum_{\sigma\in S_n/Z_n}\sum_{\tilde{\sigma}\in S_n/Z_n} 2\,\mathrm{Tr}\left(T^{a_{\sigma(1)}}...T^{a_{\sigma(n)}}\right)\; 2\,\mathrm{Tr}\left(\tilde{T}^{b_{\tilde{\sigma}(1)}}...\tilde{T}^{b_{\tilde{\sigma}(n)}}\right)\; m_n^{(0)}(\sigma,\tilde{\sigma},p).$$

The double-ordered amplitude $m_n^{(0)}(\sigma,\tilde{\sigma},p)$ is rather simple and explicitly given by

$$m_n^{(0)}(\sigma,\tilde{\sigma},p) = i(-1)^{n-3+n_{\mathrm{flip}}(\sigma,\tilde{\sigma})} \sum_{G\in\mathcal{T}_n(\sigma)\cap\mathcal{T}_n(\tilde{\sigma})} \prod_{e\in E(G)} \frac{1}{s_e}.$$

We denote by $\mathcal{T}_n(\sigma)$ the set of all ordered tree diagrams with trivalent vertices and external order σ. Two diagrams with different external orders are considered to be equivalent, if we can transform one diagram into the other by a sequence of flips. Under a flip operation one exchanges at a vertex two branches. We denote by $\mathcal{T}_n(\sigma)\cap\mathcal{T}_n(\tilde{\sigma})$ the set of diagrams compatible with the external orders σ and $\tilde{\sigma}$ and by $n_{\mathrm{flip}}(\sigma,\tilde{\sigma})$ the number of flips needed to transform any diagram from $\mathcal{T}_n(\sigma)\cap\mathcal{T}_n(\tilde{\sigma})$ with the external order σ into a diagram with the external order $\tilde{\sigma}$. The number $n_{\mathrm{flip}}(\sigma,\tilde{\sigma})$ will be the same for all diagrams from $\mathcal{T}_n(\sigma)\cap\mathcal{T}_n(\tilde{\sigma})$. For a diagram G we denote by $E(G)$ the set of the internal edges and by s_e the Lorentz invariant corresponding to the internal edge e.

Let us now consider graviton scattering amplitudes. The polarisation of an external graviton is described by a product of two spin-1 polarisation vectors

$$\varepsilon_{\mu_j\nu_j}^{\lambda_j\tilde{\lambda}_j} = \varepsilon_{\mu_j}^{\lambda_j}\varepsilon_{\nu_j}^{\tilde{\lambda}_j}.$$

We may therefore describe the polarisation configuration of n external gravitons by two n-tuples

$$\varepsilon = \left(\varepsilon_1^{\lambda_1},...,\varepsilon_n^{\lambda_n}\right), \qquad \tilde{\varepsilon} = \left(\varepsilon_1^{\tilde{\lambda}_1},...,\varepsilon_n^{\tilde{\lambda}_n}\right),$$

where for each graviton the n-tuple ε contains one polarisation vector and the n-tuple $\tilde{\varepsilon}$ the other polarisation vector. Of course, since either $(\lambda_j, \tilde{\lambda}_j) = (+,+)$ or $(\lambda_j, \tilde{\lambda}_j) = (-,-)$ we have $\varepsilon = \tilde{\varepsilon}$ for gravitons. Thus we denote the tree-level scattering amplitude for n gravitons by $\mathcal{M}_n^{(0)}(p, \varepsilon, \tilde{\varepsilon})$ It will be convenient to factor of the gravitational coupling and we define $M_n^{(0)}$ by

$$\mathcal{M}_n^{(0)}(p, \varepsilon, \tilde{\varepsilon}) = \left(\frac{\kappa}{4}\right)^{n-2} M_n^{(0)}(p, \varepsilon, \tilde{\varepsilon}).$$

We recall that there are $(n-3)!$ independent primitive tree-level amplitudes in Yang-Mills theory. Using cyclic-invariance, the Kleiss-Kuijf relations and the BCJ relations we may fix three external particles at specified positions. A basis of the independent cyclic orders is then for example given by

$$B = \{ \sigma = (\sigma_1, ..., \sigma_n) \in S_n \mid \sigma_1 = 1, \sigma_2 = 2, \sigma_n = n \}.$$

Clearly,

$$|B| = (n-3)!.$$

Let us now define a $(n-3)! \times (n-3)!$-dimensional matrix $m_{\sigma\tilde{\sigma}}$, indexed by permutations σ and $\tilde{\sigma}$ from B. We set

$$m_{\sigma\tilde{\sigma}} = m_n^{(0)}(\sigma, \tilde{\sigma}, p).$$

The entries of the matrix $m_{\sigma\tilde{\sigma}}$ are the double-ordered primitive amplitudes for the bi-adjoint scalar theory with trivalent vertices encountered in the previous paragraphs. The matrix $m_{\sigma\tilde{\sigma}}$ is invertible and we set

$$S_{\sigma\tilde{\sigma}} = \left(m^{-1}\right)_{\sigma\tilde{\sigma}}.$$

The **Kawai-Lewellen-Tye (KLT) relation** reads

$$M_n^{(0)}(p, \varepsilon, \tilde{\varepsilon}) = \sum_{\sigma, \tilde{\sigma} \in B} A_n^{(0)}(\sigma, p, \varepsilon)\; S_{\sigma\tilde{\sigma}}\, A_n^{(0)}(\tilde{\sigma}, p, \tilde{\varepsilon}),$$

where the sum runs over a basis of cyclic orders. This formula relates the n-graviton amplitude to Yang-Mills amplitudes and the bi-adjoint scalar amplitudes.