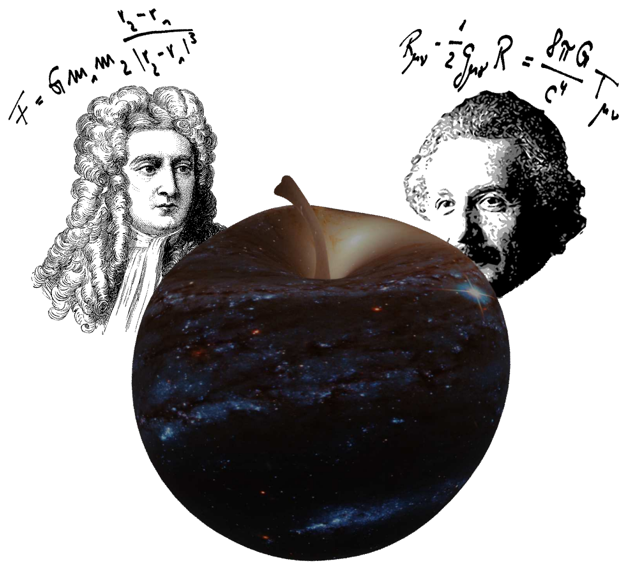
GEOMETRY IN POTSDAM

Christian Bär

General Relativity

A Mathematical Introduction



Version of March 23, 2023

Universität Potsdam

Contents

Preface

This text is based on lecture notes of various courses on General Relativity which I taught over the time. Unlike in many texts on the topic, differential geometry is not developed but assumed. There are many good introductions to this beautiful part of mathematics so that we prefer to concentrate on the actual topic - relativity. The first chapter, dealing with special relativity, does not yet assume any differential geometric background, thus giving the reader the option to learn about Lorentzian manifolds, curvature, gedesics etc. on the side.

The chapter on Special Relativity briefly recalls classical kinematics and electrodynamics emphasizing their conceptual incompatibility. It is then shown how Minkowski geometry is used to unite the two theories and to obtain what we nowadays call Special Relativity. Some famous phenomena like length contraction, time dilation, and the twin paradox are discussed. Relativistic velocity addition is investigated using hyperbolic geometry.

We then incorporate gravity by replacing Minkowski space by a more general Lorentzian manifolds whose curvature reflects the graviational field. This is made precise by Einstein's field equations and this is where differential geometry comes into play.

We then move on to the first concrete models: Robertson-Walker spacetimes are models for the whole universe on a large scale. They are used to discuss cosmic redshift, the expansion of the universe, big bang and big crunch. We also briefly discuss cosmological inflation, a phase of rapid expansion of the early universe.

The next chapter studies models for black holes, first static ones described by the Schwarzschild solution, then rotating ones modeled by the more general and more complicated Kerr solution. We discuss the trajectories of massive particles and of light and see how they differ from the classical orbits. The Kerr solution has a rich geometry and allows for time travel.

We then study gravitational waves and discuss ways to construct mathematical solutions. Finally, the Petrov classification is developed which is a way to sort spacetime models by algebraic properties of their curvature tensors.

General relativity requires a lot of computation in explicit models which can sometimes be cumbersome because of the advanced nature of differential geometry. It is reasonable to delegate this to a computer algebra system if possible. The general purpose CAS Sage has very good support for differential geometry and is used in the text in many places. It is recommended that the readers familiarize themselves with this wonderful (and free) piece of software.

It is my pleasure to thank all those who helped to improve the manuscript by suggestions, corrections or by work on the LaTeX code. My particular thanks go to Andrea Röser who wrote the first version of a part of the text in German language and created many pictures in wonderful quality, to Matthias Ludewig who translated this part of the manuscript into English and to Andreas Hermann who wrote the text for the advanced part.

Potsdam, March 2023

Christian Bär

1. Special relativity

Before starting with relativity theory we will briefly recall two older theories in physics, Newton's classical mechanics and Maxwell's electromagnetism theory. These two theories are incompatible in the sense that their laws transform differently under coordinate changes. This incompatibility was one of Einstein's main motivations to seek a theory that would combine the two. Einstein found a unification of mechanics and electromagnetism, now known under the term *special relativity theory*. In a way, Maxwell defeated Newton, as the transformation laws of special relativity are those of electrodynamics. The laws of Newtonian mechanics are only valid approximately at low velocities.

1.1. Classical kinematics

1.1.1. Absolute space

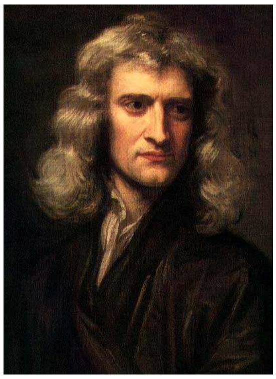
In Sir Isaac Newton's (1643-1727) world, space exists independently of all the objects contained in it. In his own words:
Absolute space, in its own nature, without regard to anything external, remains always similar and immovable.

The geometry of space is assumed to be Euclidean, i.e., it is assumed that the laws of Euclidean geometry hold for measurements performed in physical space. In other words, we can introduce Cartesian coordinates to identify space with $\mathbb{R}^3$ and then apply the usual rules of Cartesian geometry,

$$\text{absolute space} \quad \xleftrightarrow{\text{identify}} \quad \mathbb{R}^3.$$

Such a coordinate system is not unique but the Euclidean structure is invariant under coordinate transformations of the form

$$\Phi : \mathbb{R}^3 \to \mathbb{R}^3, \quad \Phi(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{b},$$



source: https://exhibitions.lib.cam.ac.uk/linesofthought/artifacts/newton-by-kneller

Figure 1.. Godfrey Kneller's portrait of Isaac Newton (1689)

with $\mathbf{A} \in \mathrm{O}(3)$, i.e., $\mathbf{A}^\top\mathbf{A} = \mathbf{I}$, and $\mathbf{b} \in \mathbb{R}^3$. The set of all such transformations is called the *Euclidean transformation group*.
It should be emphasized that such an assumption requires empirical justification. Indeed, measurements performed in every-day-life support Newton's ideas about space; if you measure the sum of angles in a triangle it will give 180 degrees to very high precision and is thus in accordance with Euclidean geometry.

1.1.2. Absolute time

Newton's ideas about time are similar to those about space:
Absolute, true and mathematical time, of itself, and from its own nature flows equably without regard to anything external.
From a mathematical point of view, this means that we can measure time by a real parameter

$$\text{absolute time} \quad \xleftrightarrow{\text{identify}} \quad \mathbb{R}.$$

More precisely, we fix a time interval, e.g., a second, and we then measure time in real multiples of this chosen time unit. The resulting identification of absolute time with $\mathbb{R}$ is unique up to transformations of the form

$$\mathbb{R} \to \mathbb{R}, \quad t \mapsto t + t_0,$$

with some fixed $t_0 \in \mathbb{R}$. Since we can distinguish future and past, we do not admit transformations of the form $\mathbb{R} \to \mathbb{R}$, $t \mapsto -t + t_0$, where $t_0 \in \mathbb{R}$.
The trajectory of a point particle is described by a curve, i.e., by a map

$$\mathbf{x} : [a, b] \to \mathbb{R}^3,$$

where to each time coordinate t we associate the corresponding space coordinates $\mathbf{x}(t) = (x^1(t), x^2(t), x^3(t))$ of the particle. Usually, we can and will assume that the curve $\mathbf{x}$ is smooth, $\mathbf{x} \in C^\infty([a, b], \mathbb{R}^3)$. The *velocity* of the particle is then given by

$$\dot{\mathbf{x}} : [a, b] \to \mathbb{R}^3$$

and the *acceleration* by

$$\ddot{\mathbf{x}} : [a, b] \to \mathbb{R}^3.$$

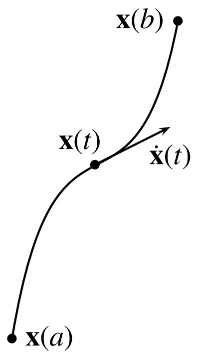


Figure 2.. Velocity vector

We measure the *mass* of the particle in real multiples of a fixed unit mass, like kilogram. Hence mass is mathematically given by a function

$$m : [a, b] \to \mathbb{R}.$$

The *momentum* is then given by

$$\mathbf{p} = m \cdot \dot{\mathbf{x}} : [a, b] \to \mathbb{R}^3$$

and the *kinetic energy* by

$$E = \frac{m \cdot \|\dot{\mathbf{x}}\|^2}{2} : [a, b] \to \mathbb{R}.$$

Finally, *the length of the trajectory* swept out by the particle can by calculated by the formula

$$\int_a^b \|\dot{\mathbf{x}}(t)\| \, dt.$$

A choice of space and time coordinates as described above will be called an *inertial frame*. According to Newton we can check whether or not our chosen coordinate system is "correct" as follows:

Newton's First Law
In any inertial frame, particles that are not subject to any force, are characterized by

$$\ddot{\mathbf{x}} = 0.$$

Note that this condition is equivalent to $\mathbf{x}(t) = \mathbf{x}(0) + t \cdot \dot{\mathbf{x}}(0)$. In other words, $\mathbf{x}$ is a straight line, parametrized at constant speed.
The transformations of space and time that were discussed above, namely Euclidean transformations and time shifts, map inertial frames to other inertial frames. Moreover, we can change an inertial frame to another one whose origin moves at constant velocity to the first one. This leads to the following set of transformations mapping inertial frames to inertial frames, the so-called *Galilean transformations*.

$$\begin{array}{ccc} \mathbb{R} \times \mathbb{R}^3 & \to & \mathbb{R} \times \mathbb{R}^3, \\ \begin{pmatrix} t \\ \mathbf{x} \end{pmatrix} & \mapsto & \begin{pmatrix} \tilde{t} \\ \tilde{\mathbf{x}} \end{pmatrix} := \begin{pmatrix} t + t_0 \\ \mathbf{A}\mathbf{x} + \mathbf{b_0} + t\mathbf{b_1} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \mathbf{b_1} & \mathbf{A} \end{pmatrix} \cdot \begin{pmatrix} t \\ \mathbf{x} \end{pmatrix} + \begin{pmatrix} t_0 \\ \mathbf{b_0} \end{pmatrix}, \end{array}$$

where $\mathbf{A} \in \mathrm{O}(3)$, $\mathbf{b_0}, \mathbf{b_1} \in \mathbb{R}^3$ and $t_0 \in \mathbb{R}$.
If $\mathbf{x}$ is the trajectory of a particle in one inertial frame, its trajectory, velocity and acceleration in another inertial frame take the form

$$\begin{aligned} \tilde{\mathbf{x}} &= \mathbf{A}\mathbf{x} + \mathbf{b_0} + t\mathbf{b_1}, \\ \dot{\tilde{\mathbf{x}}} &= \mathbf{A}\dot{\mathbf{x}} + \mathbf{b_1}, \\ \ddot{\tilde{\mathbf{x}}} &= \mathbf{A}\ddot{\mathbf{x}}. \end{aligned}$$

Observe that $\frac{d^2\mathbf{x}}{dt^2} = 0$ if and only if $\frac{d^2\tilde{\mathbf{x}}}{dt^2} = 0$ if and only if $\frac{d^2\tilde{\mathbf{x}}}{d\tilde{t}^2} = 0$. So indeed, Newton's first law is compatible with Galilean transformations. In the special case $\mathbf{A} = \mathbf{I}$ and $t_0 = 0$, we have

$-\mathbf{v} := \mathbf{b_1}$ = velocity of observer 2, measured by observer 1

$\mathbf{v_1} := \dot{\mathbf{x}}$ = velocity of the particle, measured by observer 1

$\mathbf{v_2} := \dot{\tilde{\mathbf{x}}}$ = velocity of the particle, measured by observer 2

Hence we have derived the velocity-addition formula

$$\boxed{\mathbf{v_1} = \mathbf{v} + \mathbf{v_2}} \tag{1.1}$$

Newton's second law

In any inertial system, if a particle is subject to the force $\mathbf{F}$*, then*

$$\tfrac{d}{dt}(m(t)\dot{\mathbf{x}}(t)) = \mathbf{F}(t, \mathbf{x}(t), \dot{\mathbf{x}}(t)).$$

Here a force is described by a (smooth) mapping of the form $\mathbf{F} : \mathbb{R}\times\mathbb{R}^3\times\mathbb{R}^3 \to \mathbb{R}^3$. In particular, for $m > 0$ we have

$$\ddot{\mathbf{x}}(t) = \frac{1}{m(t)}(\mathbf{F}(t, \mathbf{x}(t), \dot{\mathbf{x}}(t)) - \dot{m}(t)\dot{\mathbf{x}}(t)).$$

A solution of such an ordinary differential equation is uniquely determined by its initial values

$$\mathbf{x}(t_0) \quad \text{and} \quad \dot{\mathbf{x}}(t_0).$$

Therefore, the theory is deterministic, i.e., we can predict the future if we know the initial values.

Example 1.1. A mass m is suspended between two springs with spring constant $k > 0$. We want to find the equations of motion, given $x(0)$ and $\dot{x}(0)$.

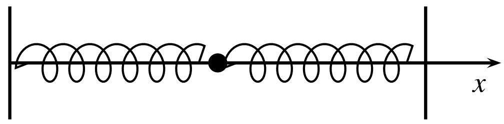


Figure 3.. Mass suspended between springs

By Hooke's law, the force is $F(t, x, y) = -kx$. From this it follows that $m\ddot{x}(t) = -kx(t)$, hence $\ddot{x}(t) = -\frac{k}{m}x(t)$. This ODE has the general solution

$$x(t) = A \cdot \sin\left(\sqrt{\tfrac{k}{m}}t\right) + B \cdot \cos\left(\sqrt{\tfrac{k}{m}}t\right),$$

where $B = x(0)$ and $\dot{x}(0) = A\sqrt{\frac{k}{m}}$. Therefore

$$x(t) = \dot{x}(0) \cdot \sqrt{\tfrac{m}{k}} \cdot \sin\left(\sqrt{\tfrac{k}{m}}t\right) + x(0) \cdot \cos\left(\sqrt{\tfrac{k}{m}}t\right).$$

Energy Equation

Let us assume that the mass m is constant. We differentiate the kinetic energy of a particle and obtain the *energy equation*

$$\frac{d}{dt}E = \frac{m}{2}\frac{d}{dt}\|\dot{\mathbf{x}}\|^2 = m\langle\ddot{\mathbf{x}},\dot{\mathbf{x}}\rangle = \langle\mathbf{F},\mathbf{v}\rangle\,. \tag{1.2}$$

1.2. Electrodynamics

Now we turn to Maxwell's electrodynamics. Assume that we are in a vacuum without any electric charges present. In this case, electric and magnetic phenomena are described by functions

$$f:\mathbb{R}\times\mathbb{R}^3\to\mathbb{R},$$

that solve the *wave equation*, i.e.,

$$\square f := \tfrac{1}{c^2}\tfrac{\partial^2 f}{\partial t^2} - \Delta f = 0,$$

where $\Delta = \sum_{i=1}^{3}\frac{\partial^2}{\partial(x^i)^2}$ is the *Laplace operator* and c the speed of light in vacuum (about $300,000$ km/s). As all observers in an inertial frame have equal right, the question arises which transformations preserve the wave equation. More precisely, which are the transformations

source: https://en.wikipedia.org/wiki/James_Clerk_Maxwell

Figure 4.. James Clerk Maxwell (1831–1879)

$$\Phi:\mathbb{R}\times\mathbb{R}^3\to\mathbb{R}\times\mathbb{R}^3,\quad \Phi(\mathbf{x}) = \mathbf{L}\mathbf{x} + \begin{pmatrix} t_0 \\ \mathbf{b_0}\end{pmatrix},$$

with $\mathbf{L}\in\mathrm{Mat}(4\times 4,\mathbb{R})$ such that whenever f solves the wave equation, so does $\tilde{f} := f\circ\Phi$? To find out, we set $x^0 := c\cdot t$ and $\mathbf{x} := (x^0, x^1, x^2, x^3)$. The wave equation then is

$$\square f = \frac{\partial^2 f}{\partial(x^0)^2} - \sum_{i=1}^{3}\frac{\partial^2 f}{\partial(x^i)^2} = 0.$$

We now calculate $\square\tilde{f}$. To this end, write

$$\mathbf{L} = (\mathbf{L_0},\mathbf{L_1},\mathbf{L_2},\mathbf{L_3}) = \begin{pmatrix} L_0^0 & L_1^0 & L_2^0 & L_3^0 \\ L_0^1 & L_1^1 & L_2^1 & L_3^1 \\ L_0^2 & L_1^2 & L_2^2 & L_3^2 \\ L_0^3 & L_1^3 & L_2^3 & L_3^3 \end{pmatrix},$$

where $\mathbf{L_i} \in \mathbb{R}^4$ for $i = 0, \dots, 3$ and $\mathbf{x_0} := (t_0, \mathbf{b_0})^\top$. We compute

$$\frac{\partial \tilde{f}}{\partial x^i} = \frac{\partial}{\partial x^i} f(\mathbf{L_0}x^0 + \mathbf{L_1}x^1 + \mathbf{L_2}x^2 + \mathbf{L_3}x^3 + \mathbf{x_0}) = \sum_{m=0}^{3} \frac{\partial f}{\partial x^m}(\mathbf{Lx} + \mathbf{x_0}) \cdot L_i^m,$$

hence

$$\frac{\partial^2 \tilde{f}}{\partial x^i \partial x^j} = \sum_{m,n=0}^{3} \frac{\partial^2 f}{\partial x^m \partial x^n}(\mathbf{Lx} + \mathbf{x_0}) L_i^m L_j^n. \tag{1.3}$$

Define

$$\mathbf{I}_{1,3} := \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

The Hessian matrix of a twice continuously differentiable function f is the symmetric matrix $\operatorname{hess} f = (\frac{\partial^2 f}{\partial x^i \partial x^j})_{i,j}$. On the vector space of real symmetric $(n \times n)$-matrices, we can define the following scalar product:

$$(\mathbf{A}, \mathbf{B})_S := \sum_{i,j=1}^{n} A^i_j B^i_j = \operatorname{tr}(\mathbf{A}^\top \cdot \mathbf{B}).$$

Then, by definition of the $\Box$-operator,

$$-\Box f = (\operatorname{hess} f, \mathbf{I}_{1,3})_S\,.$$

By (1.3) we have

$$\begin{aligned} -\Box \tilde{f} &= (\mathbf{L}^\top \cdot \operatorname{hess} f \cdot \mathbf{L}, \mathbf{I}_{1,3})_S = \operatorname{tr}(\mathbf{L}^\top \cdot \operatorname{hess} f^\top \cdot \mathbf{L} \cdot \mathbf{I}_{1,3}) \\ &= \operatorname{tr}(\operatorname{hess} f^\top \cdot \mathbf{L} \cdot \mathbf{I}_{1,3} \cdot \mathbf{L}^\top) = (\operatorname{hess} f, \mathbf{L} \cdot \mathbf{I}_{1,3} \cdot \mathbf{L}^\top)_S\,. \end{aligned}$$

We see that $\Box f = 0$ means that $\operatorname{hess} f$ is perpendicular to $\mathbf{I}_{1,3}$ while $\Box \tilde{f} = 0$ means that $\operatorname{hess} f$ is perpendicular to $\mathbf{L} \cdot \mathbf{I}_{1,3} \cdot \mathbf{L}^\top$. For these two conditions to be equivalent we must have $\mathbf{L} \cdot \mathbf{I}_{1,3} \cdot \mathbf{L}^\top = \kappa \cdot \mathbf{I}_{1,3}$ for some $\kappa \in \mathbb{R}$. Without loss of generality we will assume $\kappa = 1$ for the scaling factor, because a transformation of the form $\kappa \cdot \mathbf{I}$ just corresponds to a change of the physical unit of length.

Definition 1.2. The set of transformations

$$\mathcal{L} := \{\mathbf{L} \in \operatorname{Mat}(4 \times 4, \mathbb{R}) \,|\, \mathbf{L} \cdot \mathbf{I}_{1,3} \cdot \mathbf{L}^\top = \mathbf{I}_{1,3}\}$$

is called the *Lorentz group*. The corresponding set of affine-linear transformations

$$\mathcal{P} := \{\Phi : \mathbb{R}^4 \to \mathbb{R}^4 \,|\, \Phi(\mathbf{x}) = \mathbf{Lx} + \mathbf{x_0},\ \mathbf{L} \in \mathcal{L},\ \mathbf{x_0} \in \mathbb{R}^4\}$$

is called the *Poincaré group*.

We have seen that the "admissible" coordinate transformations of Newtonian mechanics are the Galilean transformations while those for electrodynamics are the Poincaré transformations. These two groups of transformations are not contained in one another; in this sense classical kinematics and electrodynamics are incompatible.

Now the question is: which theory is correct if any?

One should be able to answer this question by means of suitable experiments. For instance, look at the following situation: Observer 2 is located on a spacecraft that travels towards earth and sends a light signal to observer 1 on the earth. Observer 1 measures the velocity c_1 for the incoming light signal, observer 2 however measures another velocity c_2. According to (1.1), classical kinematics predicts $c_1 = c_2 + v$, where v is the velocity of the spacecraft with respect to the earth. On the other hand, the theory of electrodynamics states that the speed of light in vacuum is a fixed value c, independently of the motions of the source and the observer.

In fact, experiments, such as the famous Michelson-Morley experiment, have confirmed the predictions of electrodynamics!

source: https://en.wikipedia.org/wiki/Henri_Poincare

Figure 5.. Henri Poincaré (1854–1912)

1.3. The Lorentz group and Minkowski geometry

In order to develop a kinematic theory which is invariant under Poincaré transformations rather than Galilei transformations, we first need to understand these Poincaré transformations better. The crucial part are the Lorentz transformations because adding translations then yields all Poincaré transformations. The resulting geometry of lightlike, timelike and spacelike vectors is known as Minkowski geometry, named after the mathematician Hermann Minkowski, a close friend of David Hilbert.

source: https://en.wikipedia.org/wiki/Hermann_Minkowski

Figure 6.. Hermann Minkowski (1864 –1909)

Convention. For $(x^0, x^1, x^2, x^3) \in \mathbb{R}^4$ write $(x^0, \hat{\mathbf{x}})$ with $\hat{\mathbf{x}} := (x^1, x^2, x^3)$. We write $\langle \cdot, \cdot \rangle$ for the usual scalar product in $\mathbb{R}^n$, i.e.,

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^{n} x^i y^i.$$

We further define another inner product $\langle\!\langle \cdot, \cdot \rangle\!\rangle$ on $\mathbb{R}^4$ by

$$\langle\!\langle \mathbf{x}, \mathbf{y} \rangle\!\rangle := \langle \mathbf{x}, \mathbf{I_{1,3}} \cdot \mathbf{y} \rangle = -x^0 y^0 + \langle \hat{\mathbf{x}}, \hat{\mathbf{y}} \rangle .$$

The symmetric bilinear form $\langle\!\langle \cdot, \cdot \rangle\!\rangle$ is indefinite and non-degenerate. Recall that "non-degenerate" means that $\langle\!\langle \mathbf{x}, \mathbf{y} \rangle\!\rangle = 0$ for all $\mathbf{x} \in \mathbb{R}^4$ implies that $\mathbf{y} = 0$.

By Exercise 1.2, $\mathbf{L} \in \mathcal{L}$ if and only if $\mathbf{L}^\top \mathbf{I_{1,3}} \mathbf{L} = \mathbf{I_{1,3}}$. This is equivalent to

$$\langle \mathbf{x}, \mathbf{L}^\top \mathbf{I_{1,3}} \mathbf{L} \mathbf{y} \rangle = \langle \mathbf{x}, \mathbf{I_{1,3}} \mathbf{y} \rangle$$

for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^4$. For the left-hand-side we get $\langle \mathbf{x}, \mathbf{L}^\top \mathbf{I_{1,3}} \mathbf{L} \mathbf{y} \rangle = \langle \mathbf{L}\mathbf{x}, \mathbf{I_{1,3}} \mathbf{L} \mathbf{y} \rangle = \langle\!\langle \mathbf{L}\mathbf{x}, \mathbf{L}\mathbf{y} \rangle\!\rangle$ while the right-hand-side is $\langle \mathbf{x}, \mathbf{I_{1,3}} \mathbf{y} \rangle = \langle\!\langle \mathbf{x}, \mathbf{y} \rangle\!\rangle$. Hence we have obtained another characterization of the Lorentz group as

$$\mathcal{L} = \{\mathbf{L} \in \mathrm{Mat}(4 \times 4, \mathbb{R}) \mid \langle\!\langle \mathbf{L}\mathbf{x}, \mathbf{L}\mathbf{y} \rangle\!\rangle = \langle\!\langle \mathbf{x}, \mathbf{y} \rangle\!\rangle \text{ for all } \mathbf{x}, \mathbf{y} \in \mathbb{R}^4\}.$$

This formally resembles the definition of the orthogonal group $\mathrm{O}(n)$, which by definition is

$$\mathrm{O}(n) = \{\mathbf{A} \in \mathrm{Mat}(n \times n, \mathbb{R}) \mid \langle \mathbf{A}\mathbf{x}, \mathbf{A}\mathbf{y} \rangle = \langle \mathbf{x}, \mathbf{y} \rangle \text{ for all } \mathbf{x}, \mathbf{y} \in \mathbb{R}^n\}.$$

Definition 1.3. We call $(\mathbb{R}^4, \langle\!\langle \cdot, \cdot \rangle\!\rangle)$ the (4-dimensional) *Minkowski space*. The inner product $\langle\!\langle \cdot, \cdot \rangle\!\rangle$ is called the *Minkowski product*.

Any Lorentz transformation $\mathbf{L} \in \mathcal{L}$ has the following properties:

1) $\det(\mathbf{I_{1,3}}) = \det(\mathbf{L}^\top \mathbf{I_{1,3}} \mathbf{L}) = \det(\mathbf{L})^2 \cdot \det(\mathbf{I_{1,3}})$, hence $\det(\mathbf{L}) = \pm 1$.
2) We have

$$-1 = (\mathbf{I_{1,3}})^0_0 = (\mathbf{L}^\top \mathbf{I_{1,3}} \mathbf{L})^0_0 = -(L^0_0)^2 + \sum_{i=1}^{3} (L^0_i)^2,$$

thus

$$(L^0_0)^2 = 1 + \sum_{i=1}^{3} (L^0_i)^2. \tag{1.4}$$

In particular, we have $(L^0_0)^2 \geq 1$, i.e., $L^0_0 \geq 1$ or $L^0_0 \leq -1$.

Definition 1.4. We define the following subsets of $\mathbf{L}$:

$$\begin{aligned}
\mathcal{L}^\uparrow_+ &:= \{\mathbf{L} \in \mathcal{L} \mid \det L = +1,\ L^0_0 \geq +1\},\\
\mathcal{L}^\uparrow_- &:= \{\mathbf{L} \in \mathcal{L} \mid \det L = -1,\ L^0_0 \geq +1\},\\
\mathcal{L}^\downarrow_+ &:= \{\mathbf{L} \in \mathcal{L} \mid \det L = +1,\ L^0_0 \leq -1\},\\
\mathcal{L}^\downarrow_- &:= \{\mathbf{L} \in \mathcal{L} \mid \det L = -1,\ L^0_0 \leq -1\}.
\end{aligned}$$

The Lorentz group is the disjoint union of these four subsets, $\mathcal{L} = \mathcal{L}_+^\uparrow \sqcup \mathcal{L}_-^\uparrow \sqcup \mathcal{L}_+^\downarrow \sqcup \mathcal{L}_-^\downarrow$. In fact, one can show that they are the connected components of $\mathcal{L}$. The subset $\mathcal{L}_+^\uparrow$ is a subgroup of $\mathcal{L}$, see Exercise 1.5. The other three subsets are not because they do not contain the identity matrix. We make the further assignments for subsets of $\mathcal{L}$:

orientation preserving Lorentz tranformations:	$\mathcal{L}_+ := \mathcal{L}_+^\uparrow \sqcup \mathcal{L}_+^\downarrow$,
time orientation preserving Lorentz tranformations:	$\mathcal{L}^\uparrow := \mathcal{L}_+^\uparrow \sqcup \mathcal{L}_-^\uparrow$,
space orientation preserving Lorentz tranformations:	$\mathcal{L}_+^\uparrow \sqcup \mathcal{L}_-^\downarrow$.

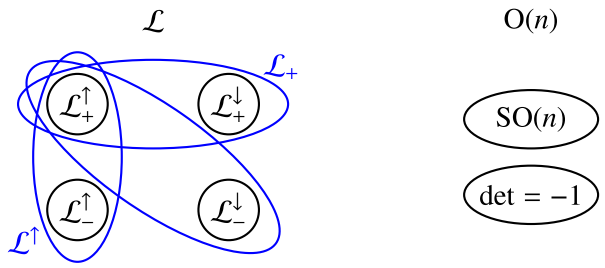


Figure 7.. Components of the Lorentz group versus orthogonal group

The Lorentz group contains the following special elements:

1) The first type consists of matrices of the form $\left(\begin{array}{c|c} 1 & 0 \\ \hline 0 & \mathbf{A} \end{array}\right)$ where $\mathbf{A} \in \mathrm{O}(3)$. For example, we have space-like rotations:

$$\underbrace{\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\varphi & -\sin\varphi & 0 \\ 0 & \sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}}_{\text{rotation about } x^3\text{-axis}} \quad \underbrace{\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\varphi & 0 & -\sin\varphi \\ 0 & 0 & 1 & 0 \\ 0 & \sin\varphi & 0 & \cos\varphi \end{pmatrix}}_{\text{rotation about } x^2\text{-axis}}$$

or space-like reflections

$$\underbrace{\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}}_{\text{reflection at } x^2\text{-}x^3\text{-plane}}$$

2) *Boosts* are Lorentz transformations which mix space and time components and leave a 2-dimensional subspace fixed; for example

$$\mathbf{L_1} = \begin{pmatrix} \cosh\varphi & \sinh\varphi & 0 & 0 \\ \sinh\varphi & \cosh\varphi & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad \text{or} \quad \mathbf{L_2} = \begin{pmatrix} \cosh\varphi & 0 & \sinh\varphi & 0 \\ 0 & 1 & 0 & 0 \\ \sinh\varphi & 0 & \cosh\varphi & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{1.5}$$

Lemma 1.5 (Hyperbolic identities). *For all $\varphi, \varphi_1, \varphi_2 \in \mathbb{R}$ we have*
(a) $\cosh(\varphi_1 + \varphi_2) = \cosh(\varphi_1)\cosh(\varphi_2) + \sinh(\varphi_1)\sinh(\varphi_2)$;
(b) $\sinh(\varphi_1 + \varphi_2) = \cosh(\varphi_1)\sinh(\varphi_2) + \sinh(\varphi_1)\cosh(\varphi_2)$;
(c) $\cosh(\varphi)^2 - \sinh(\varphi)^2 = 1$.

Proof. By definition, $\cosh\varphi = \frac{1}{2}(e^{\varphi} + e^{-\varphi})$ and $\sinh\varphi = \frac{1}{2}(e^{\varphi} - e^{-\varphi})$, hence

$$e^{\varphi} = \cosh\varphi + \sinh\varphi. \tag{1.6}$$

Inserting (1.6) into

$$\cosh(\varphi_1 + \varphi_2) = \tfrac{1}{2}(e^{\varphi_1+\varphi_2} + e^{-(\varphi_1+\varphi_2)}) = \tfrac{1}{2}(e^{\varphi_1}e^{\varphi_2} + e^{-\varphi_1}e^{-\varphi_2})$$

yields ((a)) and similarly for ((b)).
To show ((c)) we observe $\cosh\varphi - \sinh\varphi = \cosh(-\varphi) + \sinh(-\varphi) = e^{-\varphi}$. Multiplication with (1.6) gives

$$1 = e^{\varphi} \cdot e^{-\varphi} = (\cosh\varphi + \sinh\varphi)\cdot(\cosh\varphi - \sinh\varphi) = \cosh^2\varphi - \sinh^2\varphi. \qquad \square$$

Remark 1.6. Geometrically, assertion ((c)) means that for each $\varphi \in \mathbb{R}$ the point $(\cosh\varphi, \sinh\varphi)^\top$ lies on the upper branch of the hyperbola in $\mathbb{R}^2$ given by $(x^0)^2 = 1 + (x^1)^2$. In fact, $\varphi \mapsto (\cosh\varphi, \sinh\varphi)^\top$ maps $\mathbb{R}$ bijectively onto this curve.

Lemma 1.7 (Hyperbolic angular identities). *For all $\varphi_1, \varphi_2 \in \mathbb{R}$, we have*

$$\begin{pmatrix} \cosh\varphi_1 & \sinh\varphi_1 \\ \sinh\varphi_1 & \cosh\varphi_1 \end{pmatrix} \cdot \begin{pmatrix} \cosh\varphi_2 & \sinh\varphi_2 \\ \sinh\varphi_2 & \cosh\varphi_2 \end{pmatrix} = \begin{pmatrix} \cosh(\varphi_1+\varphi_2) & \sinh(\varphi_1+\varphi_2) \\ \sinh(\varphi_1+\varphi_2) & \cosh(\varphi_1+\varphi_2) \end{pmatrix}.$$

Proof. This follows directly from Lemma 1.5. $\square$

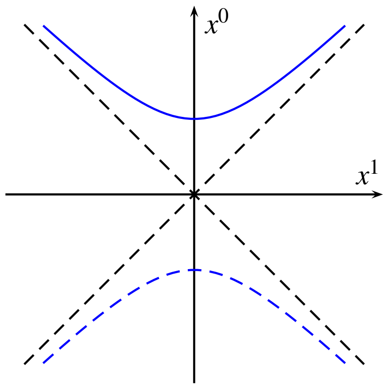


Figure 8.. Hyperbola

This lemma tells us for instance that the boosts of the form

$$\begin{pmatrix} \cosh\varphi & \sinh\varphi & 0 & 0 \\ \sinh\varphi & \cosh\varphi & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

form a subgroup of the Lorentz group. More precisely,

$$\mathbb{R} \to \mathcal{L}, \quad \varphi \mapsto \begin{pmatrix} \cosh\varphi & \sinh\varphi & 0 & 0 \\ \sinh\varphi & \cosh\varphi & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

is a group homomorphism. The first and second column traces a hyperbola when φ runs through $\mathbb{R}$.

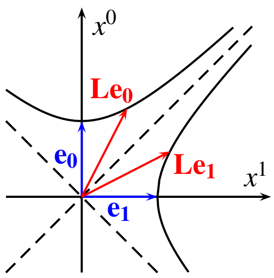


Figure 9.. Minkowski perpendicular vectors

For $\mathbf{L} = \begin{pmatrix} \cosh\varphi & \sinh\varphi \\ \sinh\varphi & \cosh\varphi \end{pmatrix}$ the vector $\mathbf{Le_0}$ is perpendicular to $\mathbf{Le_1}$ with respect to $\langle\langle\cdot,\cdot\rangle\rangle$.

Definition 1.8. A vector $\mathbf{v} \in \mathbb{R}^4$ is called

- *timelike* iff $\langle\langle \mathbf{v}, \mathbf{v} \rangle\rangle < 0$,
- *lightlike* iff $\langle\langle \mathbf{v}, \mathbf{v} \rangle\rangle = 0$ and $\mathbf{v} \neq 0$,
- *spacelike* iff $\langle\langle \mathbf{v}, \mathbf{v} \rangle\rangle > 0$ or $\mathbf{v} = 0$.

The set $C := \{\mathbf{v} \in \mathbb{R}^4 \,|\, \mathbf{v}$ lightlike $\}$ is called the *light cone*.

We observe that $\mathbf{v} = (v^0, \hat{\mathbf{v}})$ is lightlike if and only if $-(v^0)^2 + \|\hat{\mathbf{v}}\|^2 = 0$, i.e. if and only if $|v^0| = \|\hat{\mathbf{v}}\|$. This is the equation of a cone, hence the terminology "light cone".

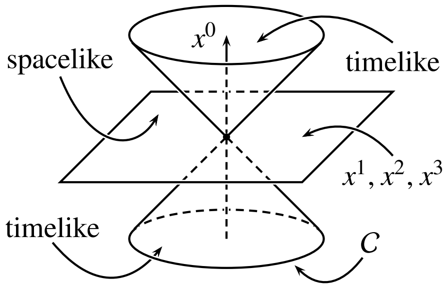


Figure 10.. Light cone and causal types

Remark 1.9. The set $\mathcal{Z} := \{\mathbf{v} \in \mathbb{R}^4 \,|\, \mathbf{v}$ timelike $\}$ is open (the "interior" of the light cone) and decomposes into two components

$$\mathcal{Z}^{\uparrow} := \{\mathbf{v} \in \mathcal{Z} \,|\, v^0 > 0\} \quad \text{and} \quad \mathcal{Z}^{\downarrow} := \{\mathbf{v} \in \mathcal{Z} \,|\, v^0 < 0\}.$$

Remark 1.10. Since $\langle\langle \mathbf{Lv}, \mathbf{Lv} \rangle\rangle = \langle\langle \mathbf{v}, \mathbf{v} \rangle\rangle$ for all vectors and Lorentz tranformations, the type (time-, light- or spacelike) of a vector $\mathbf{v} \in \mathbb{R}^4$ is left invariant under Lorentz transformations.

Whether or not a Lorentz transformation preserves $\mathcal{Z}^{\uparrow}$ and $\mathcal{Z}^{\downarrow}$ depends on the type of the Lorentz transformation. More precisely, we have

Lemma 1.11.

$$\begin{aligned}
\mathcal{L}^{\uparrow} \cdot \mathcal{Z}^{\uparrow} &= \mathcal{Z}^{\uparrow}, \\
\mathcal{L}^{\uparrow} \cdot \mathcal{Z}^{\downarrow} &= \mathcal{Z}^{\downarrow}, \\
\mathcal{L}^{\downarrow}_{\pm} \cdot \mathcal{Z}^{\uparrow} &= \mathcal{Z}^{\downarrow}, \\
\mathcal{L}^{\downarrow}_{\pm} \cdot \mathcal{Z}^{\downarrow} &= \mathcal{Z}^{\uparrow}.
\end{aligned}$$

Proof. Let $\mathbf{L} \in \mathcal{L}$ and $\mathbf{v} \in \mathcal{Z}$. We write

$$\mathbf{L} = \left(\begin{array}{c|c} L^0_0 & \mathbf{a}^\top \\ \hline \mathbf{b} & \mathbf{A} \end{array}\right) \quad \text{and} \quad \mathbf{v} = \begin{pmatrix} v^0 \\ \hat{\mathbf{v}} \end{pmatrix}$$

where $\mathbf{a}, \mathbf{b}, \hat{\mathbf{v}} \in \mathbb{R}^3$ and $\mathbf{A} \in \mathrm{Mat}(3 \times 3, \mathbb{R})$. Then

$$\mathbf{L}\mathbf{v} = \begin{pmatrix} L^0_0 v^0 + \langle \mathbf{a}, \hat{\mathbf{v}} \rangle \\ v^0 \mathbf{b} + \mathbf{A}\hat{\mathbf{v}} \end{pmatrix}.$$

We are interested in the sign of $L^0_0 v^0 + \langle \mathbf{a}, \hat{\mathbf{v}} \rangle$ because it determines whether $\mathbf{L}\mathbf{v} \in \mathcal{Z}^\uparrow$ or $\mathbf{L}\mathbf{v} \in \mathcal{Z}^\downarrow$. Equation (1.4) can be rephrased as

$$(L^0_0)^2 = 1 + \|\mathbf{a}\|^2.$$

Since $\mathbf{v}$ is timelike we have $|v^0| > \|\hat{\mathbf{v}}\|$. Therefore

$$|L^0_0 v^0|^2 = (L^0_0)^2 (v^0)^2 > (1 + \|\mathbf{a}\|^2)\|\hat{\mathbf{v}}\|^2 \geq \|\mathbf{a}\|^2 \|\hat{\mathbf{v}}\|^2 \geq \langle \mathbf{a}, \hat{\mathbf{v}} \rangle^2$$

where we used the Cauchy-Schwarz inequality in the last step. Thus $|\langle \mathbf{a}, \hat{\mathbf{v}} \rangle| < |L^0_0 v^0|$. Therefore $L^0_0 v^0 + \langle \mathbf{a}, \hat{\mathbf{v}} \rangle$ has the same sign as $L^0_0 v^0$. This shows $\mathbf{L}\mathbf{v} \in \mathcal{Z}^\uparrow$ if $L^0_0 > 0$ and $v^0 > 0$ or $L^0_0 < 0$ and $v^0 < 0$ and similarly for the other case. We have therefore shown the inclusions $\mathcal{L}^\uparrow \cdot \mathcal{Z}^\uparrow \subset \mathcal{Z}^\uparrow$, $\mathcal{L}^\uparrow \cdot \mathcal{Z}^\downarrow \subset \mathcal{Z}^\downarrow$, $\mathcal{L}^\downarrow_\pm \cdot \mathcal{Z}^\uparrow \subset \mathcal{Z}^\downarrow$, and $\mathcal{L}^\downarrow_\pm \cdot \mathcal{Z}^\downarrow \subset \mathcal{Z}^\uparrow$.
The opposite inclusions follow from the observation that if $\mathbf{L}$ lies in any of the four connected components of the Lorentz group then its inverse $\mathbf{L}^{-1}$ lies in the same component. □

Next we study "orthogonal" complements with respect to the Minkowski product. For $\mathbf{v} \in \mathbb{R}^4$ we use the notation

$$\mathbf{v}^{\perp\!\!\perp} := \{\mathbf{w} \in \mathbb{R}^4 \mid \langle\!\langle \mathbf{v}, \mathbf{w} \rangle\!\rangle = 0\}$$

to distinguish it from the usual orthogonal complement $\mathbf{v}^\perp$ which is taken with respect to the Euclidean scalar product. If $\mathbf{v} \neq 0$ then $\mathbf{v}^{\perp\!\!\perp}$ is a 3-dimensional vector subspace of $\mathbb{R}^4$.

Lemma 1.12. *Let* $\mathbf{v} \in \mathbb{R}^4$.
(a) If $\mathbf{v}$ *is timelike then all elements of* $\mathbf{v}^{\perp\!\!\perp}$ *are spacelike.*
(b) If $\mathbf{v}$ *is spacelike then* $\mathbf{v}^{\perp\!\!\perp}$ *contains timelike, lightlike and nonzero spacelike vectors.*
(c) If $\mathbf{v}$ *is lightlike then* $\mathbf{v}^{\perp\!\!\perp}$ *is the tangent space to* C *at* v *and contains lightlike and nonzero spacelike vectors but no timelike vectors.*

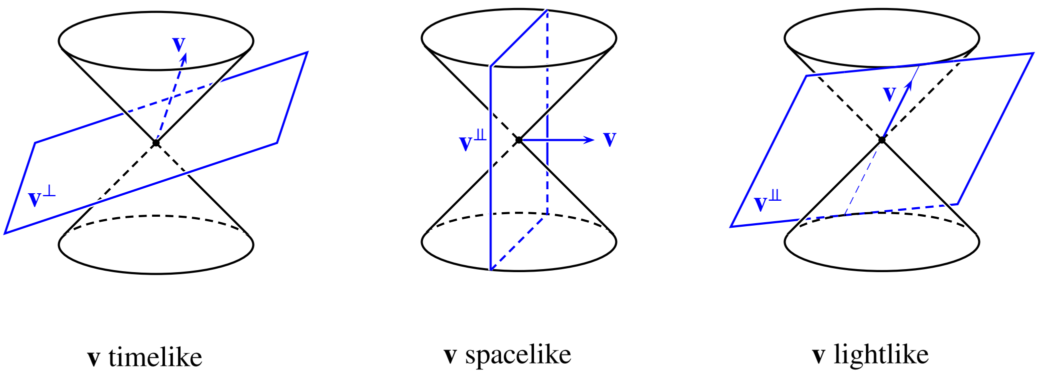


Figure 11.. Minkowski orthogonal complements

Proof. (a) Let $\mathbf{v} = \begin{pmatrix} v^0 \\ \hat{\mathbf{v}} \end{pmatrix}$ be timelike and let $\mathbf{w} \in \mathbf{v}^{⫫}$. Choose a matrix $\mathbf{A} \in \mathrm{O}(3)$ such that $\mathbf{A}\hat{\mathbf{v}} = \alpha \cdot \mathbf{e_1} = (\alpha, 0, 0)^\top$ for some $\alpha \in \mathbb{R}$. For $\mathbf{L_1} = \begin{pmatrix} 1 & 0 \\ 0 & \mathbf{A} \end{pmatrix} \in \mathcal{L}$, we have $\mathbf{L_1}\mathbf{v} = (v^0, \alpha, 0, 0)^\top$. Choose a boost $\mathbf{L_2} \in \mathcal{L}$ with

$$\mathbf{L_2} \cdot \begin{pmatrix} v^0 \\ \alpha \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} \beta \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

for some $\beta > 0$. For $\mathbf{L} = \mathbf{L_2}\mathbf{L_1} \in \mathcal{L}$ we have

$$0 = \langle\!\langle \mathbf{v}, \mathbf{w} \rangle\!\rangle = \langle\!\langle \mathbf{L}\mathbf{v}, \mathbf{L}\mathbf{w} \rangle\!\rangle = \left\langle\!\!\left\langle (\beta, 0, 0, 0)^\top, \mathbf{L}\mathbf{w} \right\rangle\!\!\right\rangle = -\beta(\mathbf{L}\mathbf{w})^0.$$

This shows $(\mathbf{L}\mathbf{w})^0 = 0$, hence $\mathbf{L}\mathbf{w} \in \{0\} \times \mathbb{R}^3$ is spacelike. Therefore $\mathbf{w} = \mathbf{L}^{-1}\mathbf{L}\mathbf{w}$ is also spacelike.
(b) Now let $\mathbf{v} \in \mathbb{R}^4$ be spacelike. If $\mathbf{v} = \mathbf{0}$ then $\mathbf{v}^{⫫} = \mathbb{R}^4$ and the statement is trivial. Hence we assume $\mathbf{v} \neq \mathbf{0}$. Since $\mathbf{v}^{⫫}$ and the x^1-x^2-x^3-hyperplane are both 3-dimensional subspaces of $\mathbb{R}^4$ their intersection has dimension 2 at least. Thus $\mathbf{v}^{⫫}$ contains nonzero spacelike vectors.
Moreover, since $\mathbf{v}$ is spacelike it is not perpendicular to itself, i.e. $\mathbf{v} \notin \mathbf{v}^{⫫}$. Hence $\mathbb{R}^4 = \mathbf{v}^{⫫} \oplus \mathbb{R} \cdot \mathbf{v}$.
We write $\mathbf{e_0} = \mathbf{w} + \alpha\mathbf{v}$ where $\mathbf{w} \in \mathbf{v}^{⫫}$ and $\alpha \in \mathbb{R}$. Then

$$-1 = \langle\!\langle \mathbf{e_0}, \mathbf{e_0} \rangle\!\rangle = \langle\!\langle \mathbf{w} + \alpha\mathbf{v}, \mathbf{w} + \alpha\mathbf{v} \rangle\!\rangle = \langle\!\langle \mathbf{w}, \mathbf{w} \rangle\!\rangle + 2\alpha \langle\!\langle \mathbf{w}, \mathbf{v} \rangle\!\rangle + \alpha^2 \langle\!\langle \mathbf{v}, \mathbf{v} \rangle\!\rangle = \langle\!\langle \mathbf{w}, \mathbf{w} \rangle\!\rangle + \alpha^2 \langle\!\langle \mathbf{v}, \mathbf{v} \rangle\!\rangle.$$

Thus

$$\langle\!\langle \mathbf{w}, \mathbf{w} \rangle\!\rangle = -1 - \alpha^2 \langle\!\langle \mathbf{v}, \mathbf{v} \rangle\!\rangle \leq -1$$

because $\mathbf{v}$ is spacelike. Thus $\mathbf{w}$ is a timelike vector in $\mathbf{v}^{⫫}$.
To find a lightlike vector in $\mathbf{v}^{⫫}$ we choose a timelike $\mathbf{w_0} \in \mathbf{v}^{⫫}$ and a nonzero spacelike $\mathbf{w_1} \in \mathbf{v}^{⫫}$. The continuous function $f(t) = \langle\!\langle t\mathbf{w_1} + (1-t)\mathbf{w_0}, t\mathbf{w_1} + (1-t)\mathbf{w_0} \rangle\!\rangle$ satisfies $f(0) < 0$ and $f(1) >$

0. Thus there exists $t_0 \in (0,1)$ with $f(t_0) = 0$. This means that $t_0\mathbf{w_1} + (1-t_0)\mathbf{w_0}$ is lightlike and since it is a linear combination of $\mathbf{w_0}$ and $\mathbf{w_1}$ it lies in $\mathbf{v}^{\perp\!\!\perp}$.
(c) Let $\mathbf{v} \in \mathbb{R}^4$ be lightlike. Let $\mathbf{w} \in T_\mathbf{v}C$. Then we can find a smooth curve $c : (-\varepsilon, \varepsilon) \to C$ with $c(0) = \mathbf{v}$ and $\dot{c}(0) = \mathbf{w}$. Since $\langle\!\langle c(t), c(t)\rangle\!\rangle = 0$ for all $t \in (-\varepsilon, \varepsilon)$ we find by differentiation

$$0 = \tfrac{d}{dt}\langle\!\langle c(t), c(t)\rangle\!\rangle|_{t=0} = \langle\!\langle \dot{c}(0), c(0)\rangle\!\rangle + \langle\!\langle c(0), \dot{c}(0)\rangle\!\rangle = 2\langle\!\langle \mathbf{v}, \mathbf{w}\rangle\!\rangle\,.$$

Thus $\mathbf{w} \in \mathbf{v}^{\perp\!\!\perp}$. This shows $T_\mathbf{v}C \subset \mathbf{v}^{\perp\!\!\perp}$. Since both space are 3-dimensional we find $T_\mathbf{v}C = \mathbf{v}^{\perp\!\!\perp}$. Intersecting with the x^1-x^2-x^3-hyperplane we see as in (b) that $\mathbf{v}^{\perp\!\!\perp}$ contains nonzero spacelike vectors. Since $\mathbf{v} \in \mathbf{v}^{\perp\!\!\perp}$ it also contains lightlike vectors. If $\mathbf{v}^{\perp\!\!\perp}$ contained a timelike vector $\mathbf{w}$ then we would have $\mathbf{v} \in \mathbf{w}^{\perp\!\!\perp}$ in contradiction to (a). □

Lemma 1.13. *Let $\mathbf{v} \in \mathbb{R}^4$ be timelike. Then*

$$\mathbf{v}^{\perp\!\!\perp} = \{\mathbf{w} \in \mathbb{R}^4 \,|\, \exists\, \alpha \in \mathbb{R}\setminus\{0\}, \textit{ such that } \alpha\mathbf{v} + \mathbf{w} \textit{ and } \alpha\mathbf{v} - \mathbf{w} \textit{ lightlike}\} \cup \{0 \in \mathbb{R}^4\}.$$

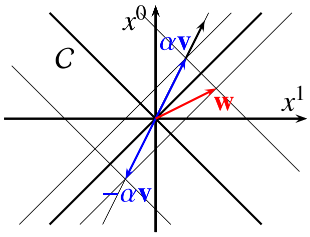


Figure 12.. Minkowski orthogonal complement of timelike vector

Proof. We show both inclusions. We start with "$\supset$": Let $\mathbf{w} \in \mathbb{R}^4$ be such that there exists $\alpha \in \mathbb{R}\setminus\{0\}$ so that $\alpha\mathbf{v} + \mathbf{w}$ and $\alpha\mathbf{v} - \mathbf{w}$ are lightlike. Then

$$0 = \langle\!\langle \alpha\mathbf{v} \pm \mathbf{w}, \alpha\mathbf{v} \pm \mathbf{w}\rangle\!\rangle = \alpha^2\langle\!\langle \mathbf{v}, \mathbf{v}\rangle\!\rangle \pm 2\alpha\langle\!\langle \mathbf{v}, \mathbf{w}\rangle\!\rangle + \langle\!\langle \mathbf{w}, \mathbf{w}\rangle\!\rangle$$

and hence

$$4\alpha\langle\!\langle \mathbf{v}, \mathbf{w}\rangle\!\rangle = 0.$$

Since $\alpha \neq 0$ this shows $\langle\!\langle \mathbf{v}, \mathbf{w}\rangle\!\rangle = 0$, i.e., $\mathbf{w} \in \mathbf{v}^{\perp\!\!\perp}$.
Now we show "$\subset$": Let $\mathbf{w} \in \mathbf{v}^{\perp\!\!\perp}\setminus\{0\}$. Then we have for all $\alpha \in \mathbb{R}$:

$$\langle\!\langle \alpha\mathbf{v} \pm \mathbf{w}, \alpha\mathbf{v} \pm \mathbf{w}\rangle\!\rangle = \alpha^2\langle\!\langle \mathbf{v}, \mathbf{v}\rangle\!\rangle \pm 2\alpha\underbrace{\langle\!\langle \mathbf{v}, \mathbf{w}\rangle\!\rangle}_{=0} + \langle\!\langle \mathbf{w}, \mathbf{w}\rangle\!\rangle = \alpha^2\langle\!\langle \mathbf{v}, \mathbf{v}\rangle\!\rangle + \langle\!\langle \mathbf{w}, \mathbf{w}\rangle\!\rangle\,.$$

Since $\langle\!\langle \mathbf{v}, \mathbf{v}\rangle\!\rangle < 0$ and $\langle\!\langle \mathbf{w}, \mathbf{w}\rangle\!\rangle \geq 0$ we can choose

$$\alpha = \sqrt{-\frac{\langle\!\langle \mathbf{w}, \mathbf{w}\rangle\!\rangle}{\langle\!\langle \mathbf{v}, \mathbf{v}\rangle\!\rangle}}$$

which does the job. □

Lemma 1.14. *Let* $\mathbf{x}, \mathbf{y} \in \mathcal{Z}^{\uparrow}$ *with* $\langle\!\langle \mathbf{x}, \mathbf{x} \rangle\!\rangle = \langle\!\langle \mathbf{y}, \mathbf{y} \rangle\!\rangle = -1$. *Then*

$$\langle\!\langle \mathbf{x}, \mathbf{y} \rangle\!\rangle \leq -1$$

and equality holds if and only if $\mathbf{x} = \mathbf{y}$.

Proof. We choose $\mathbf{A} \in O(3)$ such that $\mathbf{A}\hat{\mathbf{x}} = (\alpha, 0, 0)^{\top}$. Then we have for $\mathbf{L_1} := \begin{pmatrix} 1 & 0 \\ 0 & \mathbf{A} \end{pmatrix} \in \mathcal{L}$ that

$$\mathbf{L_1 x} = \begin{pmatrix} \beta \\ \alpha \\ 0 \\ 0 \end{pmatrix}.$$

From

$$-1 = \langle\!\langle \mathbf{x}, \mathbf{x} \rangle\!\rangle = \langle\!\langle \mathbf{L_1 x}, \mathbf{L_1 x} \rangle\!\rangle = -\beta^2 + \alpha^2$$

we see that the point $(\beta, \alpha)^{\top}$ lies on the hyperbola as in Remark 1.6. Because of $\mathbf{x} \in \mathcal{Z}^{\uparrow}$ it lies on the upper branch. Therefore there exists $\varphi \in \mathbb{R}$ such that $(\beta, \alpha)^{\top} = (\cosh\varphi, \sinh\varphi)^{\top}$. Putting $\mathbf{L_2} := \begin{pmatrix} \cosh(-\varphi) & \sinh(-\varphi) & 0 & 0 \\ \sinh(-\varphi) & \cosh(-\varphi) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \in \mathcal{L}$ and $\mathbf{L} := \mathbf{L_2 L_1} \in \mathcal{L}$ we obtain

$$\mathbf{Lx} = \mathbf{L_2 L_1 x} = \mathbf{e_0}.$$

Next we observe

$$-1 = \langle\!\langle \mathbf{y}, \mathbf{y} \rangle\!\rangle = \langle\!\langle \mathbf{Ly}, \mathbf{Ly} \rangle\!\rangle = -((\mathbf{Ly})^0)^2 + \|\widehat{\mathbf{Ly}}\|^2 \geq -((\mathbf{Ly})^0)^2$$

with equality if and only if $\widehat{\mathbf{Ly}} = \mathbf{0}$. Hence $|(\mathbf{Ly})^0| \geq 1$ with equality if and only if $\widehat{\mathbf{Ly}} = \mathbf{0}$.
Both $\mathbf{L_1}$ and $\mathbf{L_2}$ preserve time orientation, hence $\mathbf{Ly} \in \mathcal{Z}^{\uparrow}$. In other words, $(\mathbf{Ly})^0 > 0$. Therefore we know $(\mathbf{Ly})^0 \geq 1$ with equality if and only if $\widehat{\mathbf{Ly}} = \mathbf{0}$.
Now we see

$$\langle\!\langle \mathbf{x}, \mathbf{y} \rangle\!\rangle = \langle\!\langle \mathbf{Lx}, \mathbf{Ly} \rangle\!\rangle = \langle\!\langle \mathbf{e_0}, \mathbf{Ly} \rangle\!\rangle = -(\mathbf{Ly})^0 \leq -1$$

with equality if and only if $\widehat{\mathbf{Ly}} = \mathbf{0}$. Since $\mathbf{Ly} \in \mathcal{Z}^{\uparrow}$ with $\langle\!\langle \mathbf{Ly}, \mathbf{Ly} \rangle\!\rangle = -1$ the condition $\widehat{\mathbf{Ly}} = \mathbf{0}$ is equivalent to $\mathbf{Ly} = \mathbf{e_0} = \mathbf{Lx}$ and hence to $\mathbf{y} = \mathbf{x}$. □

Recall that cosh maps $[0, \infty)$ bijectively onto $[1, \infty)$. From Lemma 1.14 we know that $-\langle\!\langle \mathbf{x}, \mathbf{y} \rangle\!\rangle \in [1, \infty)$. Therefore we can make the following definition:

Definition 1.15. We set

$$H^3 := \{\mathbf{x} \in \mathcal{Z}^{\uparrow} \mid \langle\langle \mathbf{x}, \mathbf{x} \rangle\rangle = -1\}.$$

The unique function $d_H : H^3 \times H^3 \to [0, \infty)$ satisfying

$$\cosh(d_H(\mathbf{x}, \mathbf{y})) = -\langle\langle \mathbf{x}, \mathbf{y} \rangle\rangle$$

is called *hyperbolic distance*. The pair (H^3, d_H) is called the (3-dimensional) *hyperbolic space*.

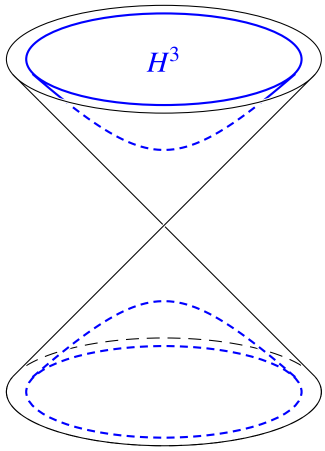


Figure 13.. Hyperbolic space

Remark 1.16. Hyperbolic space (H^3, d_H) is a metric space, i.e., for all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in H^3$, we have
(a) $d_H(\mathbf{x}, \mathbf{y}) \geq 0$ and $d_H(\mathbf{x}, \mathbf{y}) = 0$ if and only if $\mathbf{x} = \mathbf{y}$,
(b) $d_H(\mathbf{x}, \mathbf{y}) = d_H(\mathbf{y}, \mathbf{x})$,
(c) $d_H(\mathbf{x}, \mathbf{z}) \leq d_H(\mathbf{x}, \mathbf{y}) + d_H(\mathbf{y}, \mathbf{z})$.
Assertion ((b)) is clear and ((a)) is a consequence of Lemma 1.14. A proof of the triangle inequality can be found in [1, Satz 4.2.6].
Lorentz transformations which preserve the time orientation act on H^3 and preserve the hyperbolic distance. In other words, $\mathcal{L}^{\uparrow}(H^3) = H^3$ and $d_H(\mathbf{Lx}, \mathbf{Ly}) = d_H(\mathbf{x}, \mathbf{y})$ for all $\mathbf{x}, \mathbf{y} \in H^3$ and $\mathbf{L} \in \mathcal{L}^{\uparrow}$. Thus $\mathcal{L}^{\uparrow}$ acts by isometries on H^3.

Remark 1.17. For any $\mathbf{x} \in H^3$, the orthogonal complement $\mathbf{x}^{\perp\!\!\!\perp}$ coincides with the tangent space $T_{\mathbf{x}}H^3$ to H^3 at the point $\mathbf{x}$. To see this, take a smooth curve $c : (-\varepsilon, \varepsilon) \to H^3$ with $c(0) = \mathbf{x}$. Differentiating the equation

$$\langle\langle c(t), c(t) \rangle\rangle \equiv -1$$

at $t = 0$ yields

$$0 = \langle\langle \dot{c}(0), c(0) \rangle\rangle + \langle\langle c(0), \dot{c}(0) \rangle\rangle = 2 \langle\langle \dot{c}(0), \mathbf{x} \rangle\rangle$$

and hence $\dot{c}(0) \in \mathbf{x}^{\perp\!\!\!\perp}$. Thus $T_{\mathbf{x}}H^3 \subset \mathbf{x}^{\perp\!\!\!\perp}$ and since both spaces have dimension three, $T_{\mathbf{x}}H^3 = \mathbf{x}^{\perp\!\!\!\perp}$. By Lemma 1.12, $T_{\mathbf{x}}H^3$ contains only spacelike vectors. Hence the restriction of $\langle\!\langle \cdot, \cdot \rangle\!\rangle$ to $T_{\mathbf{x}}H^3$ is positive definite. Restricted to any tangent space of H^3, the Minkowski inner product becomes a Euclidean scalar product. This provides H^3 with the structure of a Riemannian manifold.

Remark 1.18. Two points $\mathbf{x}$ and $\mathbf{y} \in H^3$, $\mathbf{x} \neq \mathbf{y}$, determine a *great hyperbola* $G = G_{\mathbf{x},\mathbf{y}}$ as follows: Take the plane E that is spanned by $\mathbf{0}$, $\mathbf{x}$ and $\mathbf{y}$. The intersection of E with H^3 defines the great hyperbola $G = E \cap H^3$.

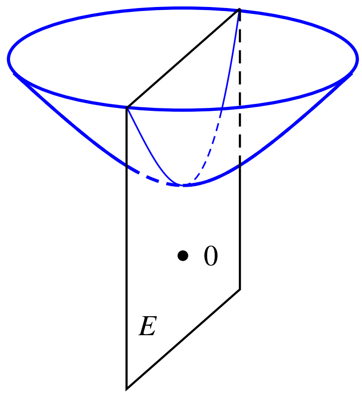


Figure 14.. Great hyperbola

We can parametrize the great hyperbola as follows: Choose $\mathbf{u} \in E \cap T_{\mathbf{x}}H^3$ with $\langle\!\langle u, u \rangle\!\rangle = 1$. The plane E contains the timelike vector $\mathbf{x}$ but $T_{\mathbf{x}}H^3 = \mathbf{x}^{\perp\!\!\!\perp}$ contains only spacelike vectors. Hence E is not contained in $T_{\mathbf{x}}H^3$ and $E \cap T_{\mathbf{x}}H^3$ must be one-dimensional. This means that there are only two possibilities to choose $\mathbf{u}$; we can only replace $\mathbf{u}$ by $-\mathbf{u}$. Both choices are equally valid. The curve parametrized by

$$c(t) = \cosh(t) \cdot \mathbf{x} + \sinh(t) \cdot \mathbf{u},$$

is contained in E, as $c(t)$ is always a linear combination of $\mathbf{x}$ and $\mathbf{u} \in E$. The curve $c(t)$ is also contained in H^3, because

$$\begin{aligned}
\langle\!\langle c(t), c(t) \rangle\!\rangle &= \langle\!\langle \cosh(t) \cdot \mathbf{x} + \sinh(t) \cdot \mathbf{u}, \cosh(t) \cdot \mathbf{x} + \sinh(t) \cdot \mathbf{u} \rangle\!\rangle \\
&= \cosh(t)^2 \underbrace{\langle\!\langle \mathbf{x}, \mathbf{x} \rangle\!\rangle}_{=-1} + 2\cosh(t)\sinh(t) \underbrace{\langle\!\langle \mathbf{x}, \mathbf{u} \rangle\!\rangle}_{=0} + \sinh(t)^2 \underbrace{\langle\!\langle \mathbf{u}, \mathbf{u} \rangle\!\rangle}_{=1} \\
&= -\cosh(t)^2 + \sinh(t)^2 \\
&= -1.
\end{aligned}$$

In fact, c passes exactly once through the great hyperbola G as t traverses the real numbers. Furthermore, $c(0) = \mathbf{x}$ and $c(\pm d_H(\mathbf{x}, \mathbf{y})) = \mathbf{y}$, where the sign depends on the choice of $\mathbf{u}$ (whether $\mathbf{u}$ points in direction of $\mathbf{y}$ or not).
Let c and $\tilde{c}$ be two great hyperbolic arcs starting at $\mathbf{x}$, parametrized by $c(t) = \cosh(t) \cdot \mathbf{x} + \sinh(t) \cdot \mathbf{u}$ and $\tilde{c}(t) = \cosh(t) \cdot \mathbf{x} + \sinh(t) \cdot \mathbf{v}$ with $\mathbf{u}, \mathbf{v} \in T_{\mathbf{x}}H^3$ and $\langle\!\langle \mathbf{u}, \mathbf{u} \rangle\!\rangle = \langle\!\langle \mathbf{v}, \mathbf{v} \rangle\!\rangle = 1$. The angle $\alpha \in [0, \pi)$

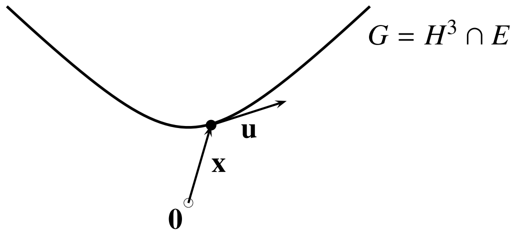


Figure 15.. Tangent vector to great hyperbola

between the two great hyperbolas is characterized by

$$\cos(\alpha) = \langle\langle \mathbf{u}, \mathbf{v} \rangle\rangle \,.$$

We have the following **trigonometric identities of hyperbolic geometry**:
Let $\mathbf{x}, \mathbf{y}, \mathbf{z} \in H^3$ three different points and let α be the angle between the great hyperbolas running from $\mathbf{x}$ to $\mathbf{y}$ and to $\mathbf{z}$, respectively. Similarly, let β be the angle at $\mathbf{y}$ and γ the angle at $\mathbf{z}$. Let furthermore $a = d_H(\mathbf{y}, \mathbf{z})$, $b = d_H(\mathbf{x}, \mathbf{z})$ and $c = d_H(\mathbf{x}, \mathbf{y})$.

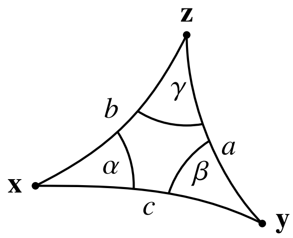


Figure 16.. Hyperbolic triangle

Theorem 1.19. *In a hyperbolic triangle the following identities hold:*
Law of sines:

$$\frac{\sinh(a)}{\sin(\alpha)} = \frac{\sinh(b)}{\sin(\beta)} = \frac{\sinh(c)}{\sin(\gamma)},$$

Law of cosines for angles:

$$\cos(\alpha) = \cosh(a)\sin(\beta)\sin(\gamma) - \cos(\beta)\cos(\gamma),$$

Law of cosines for sides:

$$\cosh(a) = \cosh(b)\cosh(c) - \sinh(b)\sinh(c)\cos(\alpha).$$

The law of cosines for sides will be helpful in the investigation of the relativistic addition of velocities. It allows to determine the length of a side of a hyperbolic triangle, given the other lengths and the opposite angle. For proofs of these laws see e.g. [2, Theorem 4.5.9].

1.4. Relativistic kinematics

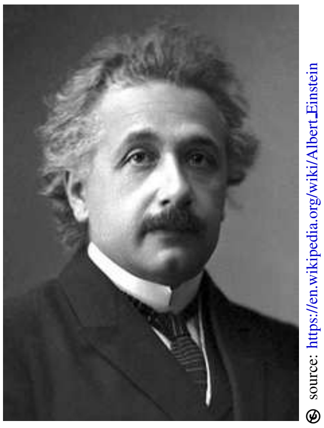


Figure 17.. Albert Einstein (1879–1955)

We now start to develop relativistic kinematics as introduced by Albert Einstein in 1905. We merge space and time to the 4-dimensional *spacetime*. The elements of spacetime are called *events*. To model particles, we use their *world lines* in spacetime

$$\{(t, \hat{\mathbf{x}}(t)) \in \mathbb{R} \times \mathbb{R}^3 \mid t \in \mathbb{R}\}.$$

instead of their parametrizations $\hat{\mathbf{x}} : \mathbb{R} \to \mathbb{R}^3$ in space. Observe that the world line is a subset of $\mathbb{R}^4$, more precisely a 1-dimensional submanifold, while $t \mapsto \hat{\mathbf{x}}(t)$ is a parametrized curve. Both contain the same information, they determine each other.

We use the canonical parametrization of the world line $t \mapsto (t, \mathbf{x}(t))$ to compute its tangent:

$$\tfrac{d}{dt}(t, \hat{\mathbf{x}}(t)) = (1, \dot{\hat{\mathbf{x}}}(t)).$$

Therefore the tangents to a world line are never parallel to $\{0\} \times \mathbb{R}^3$.

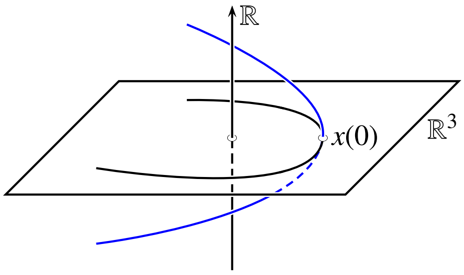


Figure 18.. Worldline

Conversely, by the implicit function theorem, any smooth curve in $\mathbb{R} \times \mathbb{R}^3$ with tangents never parallel to $\{0\} \times \mathbb{R}^3$ can be parametrized in the form

$$t \mapsto (t, \hat{\mathbf{x}}(t)).$$

Hence it is a world line.

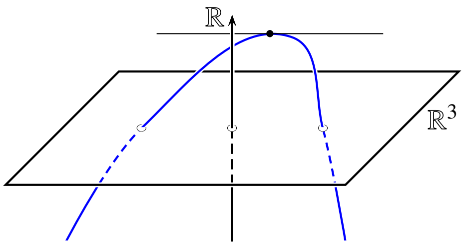


Figure 19.. Not a worldline

1.4.1. The postulate of special relativity

Inertial frames. *There exist distinguished coordinate systems for spacetime (i.e., identifications of physical spacetime with* $\mathbb{R} \times \mathbb{R}^3$*) called* inertial frames. *In an inertial frame the world lines of particles not subject to any forces are straight lines.*
Composing an inertial frame with a time orientation preserving Poincaré transformation yields another inertial frame and any two inertial frames are related by a time orientation preserving Poincaré transformation.

We compute the velocity of a particle X in an inertial frame from the view of an observer not in motion with respect to this frame, i.e., one that has the world line $\mathbb{R} \cdot \mathbf{e_0}$.

To this end, parametrize the world line of X in the form

$$t \mapsto (ct, \hat{\mathbf{x}}(t)), \quad x^0 := ct.$$

Here c is the vacuum speed of light. The *physical velocity* of the particle X is then given by

$$\mathbf{v}_{\text{phys}} := \frac{d\hat{\mathbf{x}}}{dt} = c\frac{d\hat{\mathbf{x}}}{dx^0}.$$

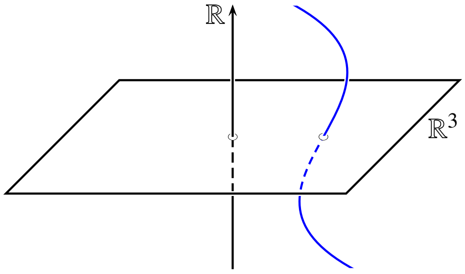


Figure 20.. Special parametrization of worldline

The *mathematical velocity* is

$$\mathbf{v} := \frac{d\hat{\mathbf{x}}}{dx^0} = \frac{1}{c}\mathbf{v}_{\text{phys}}.$$

For a reparametrization $\sigma \mapsto (\varphi(\sigma), \hat{\mathbf{x}}(\varphi(\sigma))) = \mathbf{x}(\varphi(\sigma))$ of the world line we have

$$\frac{d}{d\sigma}(\mathbf{x} \circ \varphi) = q\Big(\varphi'(\sigma), \frac{d\hat{\mathbf{x}}}{dx^0}(\varphi(\sigma)) \cdot \varphi'(\sigma)\Big) = \varphi'(\sigma)\Big(1, \frac{d\hat{\mathbf{x}}}{dx^0}(\varphi(\sigma))\Big).$$

This implies the invariance of the mathematical velocity under reparametrizations:

$$\mathbf{v} = \frac{d\hat{\mathbf{x}}}{dx^0} = \frac{d(\hat{\mathbf{x}} \circ \varphi)}{d\sigma} \Big/ \frac{d(x^0 \circ \varphi)}{d\sigma}.$$

The mathematical velocity of the particle X is determined by the slope of the tangent: For a tangent vector $\dot{\mathbf{x}}(s) = (\dot{x}^0(s), \dot{\hat{\mathbf{x}}}(s))$, we have

$$\begin{aligned}
\langle\!\langle \dot{\mathbf{x}}(s), \dot{\mathbf{x}}(s) \rangle\!\rangle &= \Big\langle\!\!\Big\langle (\dot{x}^0(s), \dot{\hat{\mathbf{x}}}), (\dot{x}^0(s), \dot{\hat{\mathbf{x}}}) \Big\rangle\!\!\Big\rangle \\
&= -\dot{x}^0(s)^2 + \left\|\dot{\hat{\mathbf{x}}}(s)\right\|^2 \\
&= \dot{x}^0(s)^2(-1 + |\mathbf{v}(s)|^2).
\end{aligned}$$

We observe:

$$\begin{array}{llllllll}
\dot{\mathbf{x}}(s) \text{ is timelike} & \Leftrightarrow & -1 + |\mathbf{v}(s)|^2 < 0 & \Leftrightarrow & |\mathbf{v}(s)| < 1 & \Leftrightarrow & \left\|\mathbf{v}_{\text{phys}}(s)\right\| < c, \\
\dot{\mathbf{x}}(s) \text{ is lightlike} & \Leftrightarrow & -1 + |\mathbf{v}(s)|^2 = 0 & \Leftrightarrow & |\mathbf{v}(s)| = 1 & \Leftrightarrow & \left\|\mathbf{v}_{\text{phys}}(s)\right\| = c, \\
\dot{\mathbf{x}}(s) \text{ is spacelike} & \Leftrightarrow & -1 + |\mathbf{v}(s)|^2 > 0 & \Leftrightarrow & |\mathbf{v}(s)| > 1 & \Leftrightarrow & \left\|\mathbf{v}_{\text{phys}}(s)\right\| > c.
\end{array}$$

We measured velocity of X with respect to an observer B1 with world line $\mathbb{R}\cdot\mathbf{e_0}$ in a given inertial frame. Which velocity of X will be measured by a second observer B2 moving with constant velocity (less than c) with respect to B1?
The world line of B2 is a straight line in the inertial frame of B1. This straight line is timelike because the velocity of B2 with respect to B1 is less than c. Thus we can find a time orientation

preserving Poincaré transformation P which maps the world line of B2 to $\mathbb{R}\mathbf{e_0}$. By the postulate of special relativity, this yields another inertial frame. We write $P(\mathbf{x}) = \mathbf{L}\mathbf{x} + \mathbf{p}$ with $\mathbf{L} \in \mathcal{L}^{\uparrow}$ and $\mathbf{p} \in \mathbb{R}^4$. In the new coordinates, B2 has the world line $\mathbb{R}\mathbf{e_0}$ and the world line of X is parametrized by

$$t \mapsto \mathbf{y}(t) := P(ct, \hat{\mathbf{x}}(t)) = P(\mathbf{x}(t)) = \mathbf{L}\mathbf{x}(t) + \mathbf{p}.$$

Thus B2 observes the tangent vector $\dot{\mathbf{y}}(t) = \mathbf{L}\dot{\mathbf{x}}(t)$ to the world line of X and measures the mathematical velocity $\frac{\hat{\dot{\mathbf{y}}}(t)}{\dot{y}^0(t)}$. Here we have $\dot{\mathbf{y}} = \dot{y}^0 \cdot \mathbf{e_0} + \hat{\dot{\mathbf{y}}}$, the splitting of $\dot{\mathbf{y}}$ into the part tangential to e_0 and the normal part corresponding to the factorization $\mathbb{R}^4 = \mathbb{R}\mathbf{e_0} \oplus \mathbf{e_0}^{\perp\!\!\!\perp} = \mathbb{R}\mathbf{e_0} \oplus T_{\mathbf{e_0}}H^3$. Reversing the transformation yields

$$\begin{aligned}
\dot{\mathbf{x}} &= \mathbf{L}^{-1}\dot{\mathbf{y}} \\
&= \mathbf{L}^{-1}(\dot{y}^0 \cdot \mathbf{e_0} + \hat{\dot{\mathbf{y}}}) \\
&= \dot{y}^0 \cdot \mathbf{L}^{-1}\mathbf{e_0} + \mathbf{L}^{-1}\hat{\dot{\mathbf{y}}} \\
&= -\langle\!\langle \dot{\mathbf{y}}, \mathbf{e_0} \rangle\!\rangle \cdot \mathbf{L}^{-1}\mathbf{e_0} + \mathbf{L}^{-1}\hat{\dot{\mathbf{y}}} \\
&= -\left\langle\!\!\left\langle \dot{\mathbf{x}}, \mathbf{L}^{-1}\mathbf{e_0} \right\rangle\!\!\right\rangle \cdot \mathbf{L}^{-1}\mathbf{e_0} + \mathbf{L}^{-1}\hat{\dot{\mathbf{y}}} \\
&= -\langle\!\langle \dot{\mathbf{x}}, \mathbf{z} \rangle\!\rangle \cdot \mathbf{z} + \mathbf{L}^{-1}\hat{\dot{\mathbf{y}}}
\end{aligned}$$

where we put $\mathbf{z} := \mathbf{L}^{-1}\mathbf{e_0} \in H^3$ for the tangent vector to the world line of B2 (with respect to the original inertial frame before transformation). Thus the mathematical velocity vector observed by B2 but expressed in the inertial frame of B1 is given by

$$\frac{\mathbf{L}^{-1}\hat{\dot{\mathbf{y}}}}{-\langle\!\langle \dot{\mathbf{x}}, \mathbf{z} \rangle\!\rangle} = \frac{\dot{\mathbf{x}} + \langle\!\langle \dot{\mathbf{x}}, \mathbf{z} \rangle\!\rangle \cdot \mathbf{z}}{-\langle\!\langle \dot{\mathbf{x}}, \mathbf{z} \rangle\!\rangle} \in \mathbf{z}^{\perp\!\!\!\perp} = T_{\mathbf{z}}H^3. \tag{1.7}$$

1.4.2. Simultaneity

How does our inertial observer B1 determine if two events happen simultaneously?
Observer B1 sends out a light signal at the event $-\alpha v$, which is reflected at the event E2 and is again received by B1 at the event αv. Because the light took the same time for the way to E2 and the way back, the event E2 must happen at the same time as the event E1= $\mathbf{0}$. For observer B2 having constant velocity with respect to B1, the event E1 happens at the same time as E2'. But for B1, the events E1 and E2' do *not* happen simultaneously.

Lemma 1.13 can now be interpreted as follows: The set of events that are simultaneous to $\mathbf{0} \in \mathbb{R} \times \mathbb{R}^3$ for an observer with world line $\mathbb{R}\mathbf{v}$, is precisely $\mathbf{v}^{\perp\!\!\perp}$.
There is a different way to establish this statement: For observer B1, two events are simultaneous if and only if they have the same x^0 component, as the x^0 component was introduced to by the time component from the view of B1.
The hyperplanes $\{x^0\} \times \mathbb{R}^3$ in $\mathbb{R}^4$ with fixed x^0 are exactly those perpendicular to the world line of B1 (with respect to the Mikowski product). Using a Lorentz transformation that converts B1 to B2, this converts events which are simultaneous for B1 into events which are simultaneous for B2, by the postulate of special relativity. On the other hand, we know that our Lorentz transformation respects the Minkowski product, in particular, it maps $\mathbf{e_0}^{\perp\!\!\perp}$ to $\mathbf{v}^{\perp\!\!\perp}$.

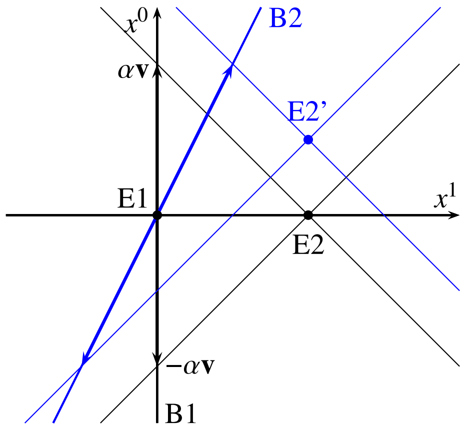


Figure 21.. Event E2 is simultaneous to E1 w.r.t. observer B1 and E2' w.r.t. B2

This shows that simultaneity of events is seen differently by different inertial observers. But who is right? Since no inertial observer is distinguished amongst all of them, every inertial observer is equally right. We have to abandon the idea that simultaneity of two events is a property of events only; simultaneity is not an *absolute concept*. Simultaneity is a *relative* concept in the sense that it depends on the observer.

Remark 1.20. "Being in the same place" is already a relative concept in classical mechanics. If observer B2 is moving with constant velocity with respect to the inertial observer B1, then, from the point of view of B1, B2 occupies different locations at different times, while B2 considers himself as staying in the same place for all times.

1.4.3. Superluminal velocity

Consider the world line $\mathbb{R}\mathbf{w}$ of a hypothetical particle X moving with constant speed higher than that of light with respect to an inertial observer B1. In other words, the tangent vector $\mathbf{w}$ to the world line of X is spacelike. The vector $\mathbf{w}$ is also nonzero because otherwise it would not span a worldline. By Lemma 1.12 ((b)) we can find a timelike vector $\mathbf{v}$ in $\mathbf{w}^{\perp\!\!\!\perp}$. Let now B2 be the inertial observer with the world line $\mathbb{R}\mathbf{v}$.

Then the whole world line of $\mathbb{R}\mathbf{w}$ is perpendicular to the world line of B2, i.e., from the point of view of B2 all events on $\mathbb{R}\mathbf{w}$ are simultaneous. Thus B2 observes the particle X to be at all places at the same time. Nothing like this has ever been observed. In B2's inertial system the "worldline" of X is not a worldline any more.

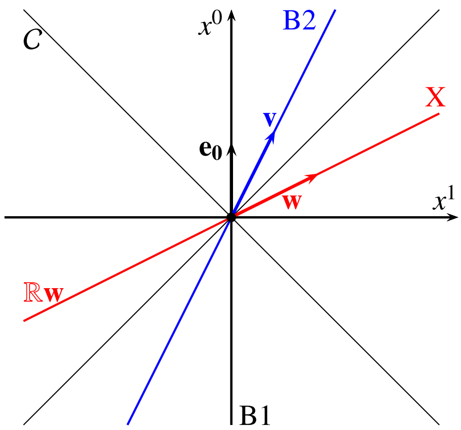


Figure 22.. Problems with superluminal velocity

Even worse: We can choose $\mathbf{u} \in \mathcal{Z}^{\uparrow}$ (close enough to the lightcone) such that $\mathbf{w}$ lies below $\mathbf{u}^{\perp\!\!\!\perp}$. Denote the observer with worldline $\mathbb{R}\mathbf{u}$ by B3. If we now switch to B3's inertial system by applying a Lorentz transformation $\mathbf{L} \in \mathcal{L}^{\uparrow}$ with $\mathbf{L}\mathbf{u} = \mathbf{e_0}$, then $\mathbf{L}\mathbf{w}$ will have negative x^0-component. From the point of view of B3, the world line of X moves into the past!

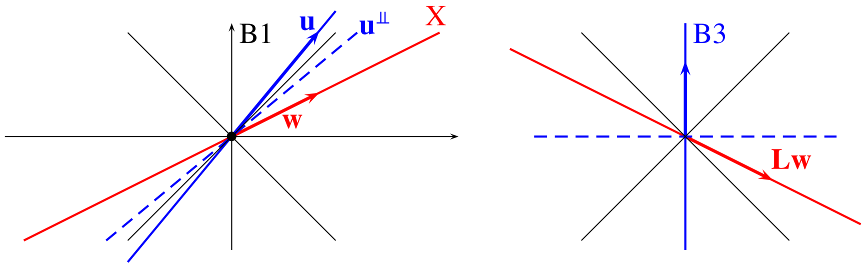


Figure 23.. Causality problems due to superluminal velocity

This leads to causality problems. If one were able to travel to or send signals to the past one could influence the past and thus change the present. These considerations lead to the conclusion that nothing can move faster than light. Hypothetical particles moving at superluminous velocity are sometimes called *tachyons*.

1.4.4. Absolute velocity and hyperbolic distance

Let X be a particle with world line $\mathbb{R}\mathbf{x} + \mathbf{p}$ and B an inertial observer with world line $\mathbb{R}\mathbf{y} + \mathbf{q}$. Here $\mathbf{p}, \mathbf{q} \in \mathbb{R}^4$ are arbitrary vectors and, without loss of generality, we can normalize $\mathbf{x}$ and $\mathbf{y}$ such that $\mathbf{x}, \mathbf{y} \in H^3$. By (1.7) B observes the particle X to have the mathematical velocity

$\mathbf{v} = \frac{\mathbf{x}+\langle\!\langle \mathbf{x},\mathbf{y}\rangle\!\rangle\cdot\mathbf{y}}{-\langle\!\langle \mathbf{x},\mathbf{y}\rangle\!\rangle}$. For the square of the absolute velocity, we calculate

$$\begin{aligned}
\langle\!\langle \mathbf{v},\mathbf{v}\rangle\!\rangle &= \frac{\overbrace{\langle\!\langle \mathbf{x},\mathbf{x}\rangle\!\rangle}^{=-1} + 2\langle\!\langle \mathbf{x},\mathbf{y}\rangle\!\rangle^2 + \langle\!\langle \mathbf{x},\mathbf{y}\rangle\!\rangle^2 \overbrace{\langle\!\langle \mathbf{y},\mathbf{y}\rangle\!\rangle}^{=-1}}{\langle\!\langle \mathbf{x},\mathbf{y}\rangle\!\rangle^2} \\
&= \frac{-1+\langle\!\langle \mathbf{x},\mathbf{y}\rangle\!\rangle^2}{\langle\!\langle \mathbf{x},\mathbf{y}\rangle\!\rangle^2} \\
&= \frac{-1+\cosh(d_H(\mathbf{x},\mathbf{y}))^2}{\cosh(d_H(\mathbf{x},\mathbf{y}))^2} \\
&= \frac{\sinh(d_H(\mathbf{x},\mathbf{y}))^2}{\cosh(d_H(\mathbf{x},\mathbf{y}))^2} \\
&= \tanh(d_H(\mathbf{x},\mathbf{y}))^2.
\end{aligned}$$

Therefore we get

$$\boxed{|\mathbf{v}| = \tanh(d_H(\mathbf{x},\mathbf{y}))}$$

for the absolute mathematical velocity.

1.4.5. Addition of velocity

As preparation we need a little lemma on hyperbolic functions.

Lemma 1.21. *For any $x \in (-1,1)$ we have*

(a) $\cosh(\operatorname{artanh}(x)) = \dfrac{1}{\sqrt{1-x^2}}$;

(b) $\sinh(\operatorname{artanh}(x)) = \dfrac{x}{\sqrt{1-x^2}}$;

(c) $e^{\operatorname{artanh}(x)} = \dfrac{1+x}{\sqrt{1-x^2}} = \dfrac{\sqrt{1+x}}{\sqrt{1-x}}$.

Proof. (a) Set $y := \operatorname{artanh}(x)$. Then

$$x^2 = \tanh(y)^2 = \frac{\sinh(y)^2}{\cosh(y)^2} = \frac{\cosh(y)^2-1}{\cosh(y)^2} = 1 - \frac{1}{\cosh(y)^2},$$

which implies $\cosh(y)^2 = \frac{1}{1-x^2}$. Because cosh is positive, we are allowed to take the positive square root, which gives the statement.

(b) From (a), we get

$$\sinh(\operatorname{artanh}(x))^2 = \cosh(\operatorname{artanh}(x))^2 - 1 = \frac{1}{1-x^2} - 1 = \frac{x^2}{1-x^2}.$$

Here we have to be careful with the sign, namely we have

$$x > 0 \Leftrightarrow \operatorname{artanh}(x) > 0 \Leftrightarrow \sinh(\operatorname{artanh}(x)) > 0.$$

Taking the square root with the correct sign yields the claim.
(c) follows from $e^y = \cosh(y) + \sinh(y)$. □

Let us now consider the following situation: We have an inertial observer with world line $\mathbb{R}\mathbf{x}+\mathbf{p}$, and inertial observer B2 with world line $\mathbb{R}\mathbf{y}+\mathbf{q}$ and an object X with world line $\mathbb{R}\mathbf{z}+\mathbf{r}$. Let $v = \tanh(d_H(\mathbf{y},\mathbf{z}))$ the absolute velocity of X in the view of B2 and $V = \tanh(d_H(\mathbf{x},\mathbf{y}))$ the absolute velocity of B2 in the view of B1. We want to determine the absolute velocity $w = \tanh(d_H(\mathbf{x},\mathbf{z}))$ of X in the view of B1. This is a problem of hyperbolic trigonometry.

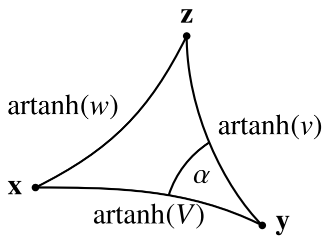


Figure 24.. Addition of velocities and hyperbolic trigonometry

Write α for the angle at the vertex $\mathbf{y}$ in this hyperbolic triangle. This is the angle between the two velocity vectors of B1 and of X in the view of B2. The law of cosines for sides of the hyperbolic geometry (Theorem 1.19) now states

$$\cosh(\operatorname{artanh}(w)) = \cosh(\operatorname{artanh}(v))\cosh(\operatorname{artanh}(V)) - \sinh(\operatorname{artanh}(v))\sinh(\operatorname{artanh}(V))\cos(\alpha)$$

and Lemma 1.21 gives

$$\begin{aligned}\frac{1}{\sqrt{1-w^2}} &= \frac{1}{\sqrt{1-v^2}}\frac{1}{\sqrt{1-V^2}} - \frac{v}{\sqrt{1-v^2}}\frac{V}{\sqrt{1-V^2}}\cos(\alpha)\\ &= \frac{1-vV\cos(\alpha)}{\sqrt{(1-v^2)(1-V^2)}}.\end{aligned}$$

Hence

$$1-w^2 = \frac{(1-v^2)(1-V^2)}{(1-vV\cos(\alpha))^2}$$

and therefore

$$\begin{aligned}w^2 &= 1-\frac{(1-v^2)(1-V^2)}{(1-vV\cos(\alpha))^2}\\ &= \frac{(1-vV\cos(\alpha))^2-(1-v^2)(1-V^2)}{(1-vV\cos(\alpha))^2}\\ &= \frac{1-2vV\cos(\alpha)+v^2V^2\cos(\alpha)^2-(1-v^2-V^2+v^2V^2)}{(1-vV\cos(\alpha))^2}\\ &= \frac{v^2+V^2-2vV\cos(\alpha)-v^2V^2\sin(\alpha)^2}{(1-vV\cos(\alpha))^2}.\end{aligned}$$

This gives the general formula for relativistic addition of velocities:

$$\boxed{w = \frac{\sqrt{v^2 + V^2 - 2vV\cos(\alpha) - v^2V^2\sin(\alpha)^2}}{1 - vV\cos(\alpha)}}$$

Let us look at two special cases. For $\alpha = \pi$, we have $\cos(\alpha) = -1$ and $\sin(\alpha) = 0$. Hence we get

$$w = \frac{\sqrt{v^2 + V^2 + 2vV}}{1 + vV} = \frac{v + V}{1 + vV}.$$

The is a deviation from the classical result $w = v + V$ by the factor $\frac{1}{1+vV}$. For velocities that are small compared to the speed of light, vV is very small and the difference is barely measurable. Now look at the case that the velocities are perpendicular to each other. For $\alpha = \pi/2$, we have $\cos(\alpha) = 0$ and $\sin(\alpha) = 1$. Therefore we get

$$w = \sqrt{v^2 + V^2 - v^2V^2}.$$

In classical mechanics, the Pythagorean theorem would have given the result $w = \sqrt{v^2 + V^2}$. For general α the law of cosines for the Euclidean geometry yields

$$w = \sqrt{v^2 + V^2 - 2vV\cos(\alpha)}$$

for classical mechanics.

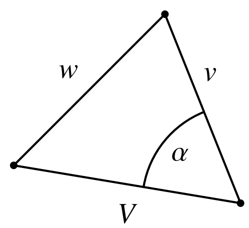


Figure 25.. Newtonian addition of velocities and Euclidean trigonometry

It is also interesting to consider that case $v = 1$, i.e., X moves with the speed of light relative to B2. Relativistic velocity addition gives us

$$w = \frac{\sqrt{1 + V^2 - 2V\cos(\alpha) - V^2\sin(\alpha)^2}}{1 - V\cos(\alpha)} = \frac{\sqrt{1 + V^2\cos(\alpha)^2 - 2V\cos(\alpha)}}{1 - V\cos(\alpha)} = 1.$$

Thus X also moves with the same speed of light relative of B1, independently of the relative motion of B1 and B2.

1.4.6. Length contraction

Consider a bar not subject to any acceleration. We choose the coordinate system such that one end of the bar has the world line $\mathbb{R}\mathbf{e_0}$ and the other end has the world line $\mathbb{R}\mathbf{e_0} + L\mathbf{e_1}$. An inertial observer B1 sitting at the first end of the bar, i.e. having world the world line $\mathbb{R}\mathbf{e_0}$, measures L for the length of the bar. Indeed, from B1's point of view, the events $\mathbf{0}$ and $(0, L, 0, 0)^\top$ are simultaneous their distance in space is

$$\sqrt{\langle\langle \mathbf{0} - (0, L, 0, 0)^\top, \mathbf{0} - (0, L, 0, 0)^\top \rangle\rangle} = \sqrt{\langle\langle (0, L, 0, 0)^\top, (0, L, 0, 0)^\top \rangle\rangle} = L.$$

Let now B2 be a second inertial observer with world line $\mathbb{R}\mathbf{x}$. Which length $\tilde{L}$ will by measured by B2?

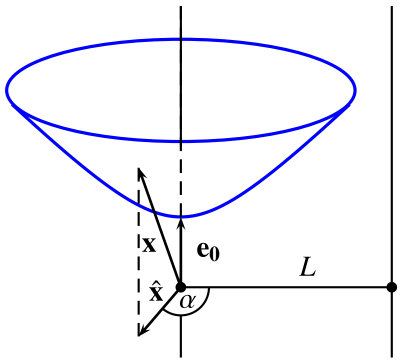


Figure 26.. Length contraction

To calculate this, we have to determine the event on the world line $\mathbb{R}\mathbf{e_0} + L\mathbf{e_1}$ that is simultaneous to $\mathbf{0}$. From B2's point of view, the events simultaneous to $\mathbf{0}$ are exactly the points on $\mathbf{x}^{\perp\!\!\!\perp}$. We solve

$$0 = \left\langle\!\!\left\langle (t, L, 0, 0)^\top, \mathbf{x} \right\rangle\!\!\right\rangle = -tx^0 + Lx^1$$

for t and we obtain

$$t = L\frac{x^1}{x^0} = L\frac{\langle \hat{\mathbf{x}}, \mathbf{e_1} \rangle}{x^0} = L\frac{\cos(\alpha) \cdot |\hat{\mathbf{x}}|}{x^0} = L \cdot \cos(\alpha) \cdot V,$$

where V is the absolute velocity between B1 and B2 and α is the angle between $\mathbf{e_1}$ and the velocity vector. Hence for B2, the events $\mathbf{0}$ and $(L \cdot \cos(\alpha) \cdot V, L, 0, 0)^\top$ are simultaneous. B2 measures the distance in space

$$\begin{aligned}
\tilde{L}^2 &= \left\langle\!\!\left\langle \mathbf{0} - (L \cdot \cos(\alpha) \cdot V, L, 0, 0)^\top, \mathbf{0} - (L \cdot \cos(\alpha) \cdot V, L, 0, 0)^\top \right\rangle\!\!\right\rangle \\
&= \left\langle\!\!\left\langle (L \cdot \cos(\alpha) \cdot V, L, 0, 0)^\top, (L \cdot \cos(\alpha) \cdot V, L, 0, 0)^\top \right\rangle\!\!\right\rangle \\
&= -L^2 \cdot \cos(\alpha)^2 \cdot V^2 + L^2 \\
&= L^2 \cdot (1 - \cos(\alpha)^2 V^2)
\end{aligned}$$

and therefore

$$\boxed{\tilde{L} = L \cdot \sqrt{1 - \cos(\alpha)^2 V^2}}$$

An inertial observer moving relative to the bar observes a length which is shortened by the factor $\sqrt{1 - \cos(\alpha)^2 V^2} \leq 1$ compared to an observer who is at rest relative to the bar. This phenomenon is known as *length contraction*.

If the motion of the two observers is in the direction of the bar ($\alpha = 0$ or $\alpha = \pi$), then we have the strongest length contraction, namely by the factor $\sqrt{1-V^2}$. If the motion is perpendicular to the bar ($\alpha = \pm\pi/2$), then there is no contraction.
The length of an object measured by an observer in rest relatively to the object is called the *proper length* or *rest length* of the object. It is the maximal length of the object that an observer can measure.
Hence the length of an object has also become a relative concept in the sense that it depends on the observer. This leads to a number of puzzling questions. Here is an example:

The Tunnel Paradox. A train with proper length L is travelling through a tunnel that also has proper length L. Is the train contained completely in the tunnel at some point?
From an outside view:
Because of length contraction, the train is shorter than the tunnel. Therefore, for some time, the train is completely contained in the tunnel.
From the locomotive driver's point of view:
Because of length contraction, the tunnel is shorter than the train. Therefore the train is never completely contained in the tunnel.
Who is right?
"Being completely contained in the tunnel" means that both ends of the train are in the tunnel *simultaneously*. Simultaneity, however, is a relative concept (depending on the inertial observer) and hence this is also the case for the concept of "being completely contained in the tunnel". Both observers are right from there respective points of view.

1.4.7. Time dilation

An inertial observer B1 with world line $\mathbb{R}\cdot\mathbf{e_0}$ observes the elapsed time T between the events $\mathbf{0}$ and $T\cdot\mathbf{e_0}$. More generally, if B1 has the world line $\mathbb{R}\cdot\mathbf{x}$ with $\mathbf{x}\in H^3$, then B1 observes the elapsed time T between the events $\mathbf{0}$ and $T\cdot\mathbf{x}$. Let now B2 be another inertial observer with world line $\mathbb{R}\cdot\mathbf{y}$ where $\mathbf{y}\in H^3$. Which is the time $\tilde{T}$ elapsed between the events $\mathbf{0}$ und $T\cdot\mathbf{x}$ from the viewpoint of B2?
To answer this question, we have to determine the $\tilde{T}$ for which the event $\tilde{T}\cdot\mathbf{y}$ is simultaneous to the event $T\cdot\mathbf{x}$ from the point of view of B2. This is the case if and only if the difference vector $T\cdot\mathbf{x}-\tilde{T}\cdot\mathbf{y}$ is Minkowski-perpendicular to the world line of B2, i.e. to $\mathbf{y}$. We solve:

$$0 = \left\langle\!\!\left\langle \mathbf{y}, T\cdot\mathbf{x}-\tilde{T}\cdot\mathbf{y}\right\rangle\!\!\right\rangle = T\cdot\langle\!\langle \mathbf{y},\mathbf{x}\rangle\!\rangle - \tilde{T}\cdot\langle\!\langle \mathbf{y},\mathbf{y}\rangle\!\rangle = -T\cdot\cosh(d_H(\mathbf{y},\mathbf{x}))+\tilde{T},$$

hence, by Lemma 1.21,

$$\tilde{T} = T\cdot\cosh(d_H(\mathbf{y},\mathbf{x})) = T\cdot\cosh(\operatorname{artanh}(V)) = T\cdot\frac{1}{\sqrt{1-V^2}},$$

where V is the velocity between B1 and B2. Because of the correction factor $\frac{1}{\sqrt{1-V^2}}\geq 1$, the time elapsed is longer in the view of B2. This phenomenon is known as *time dilation*. From the point of view of B2, the clock of B1 runs slower than his own. Exchanging roles of B1 and B2, we analogously obtain that the clock of B2 runs slower than his own for B1.

In physical units, we have

$$\boxed{\tilde{T} = \frac{T}{\sqrt{1 - \frac{V_{\text{phys}}^2}{c^2}}}}$$

For velocities much below the speed of light, $V_{phys} \ll c$, i.e. $V \ll 1$, the correction factor is very close to 1. For this reason, time dilation is not noticed in daily life.

Example 1.22. Cosmic radiation generates certain elementary particles, so-called muons, on impact with the outer atmosphere. These muons have a mean lifetime of $2 \cdot 10^{-6}$ s. Even at the speed of light, the muons can cover a distance of only

$$3 \cdot 10^5 \frac{km}{s} \cdot 2 \cdot 10^{-6} s = 600m$$

on average. One would expect that only very few muons ever reach the surface of the earth because the distance between the outer atmosphere and the surface of the earth is roughly 10 km. It is a fact however, that muons can be detected on the earth's surface in great numbers.

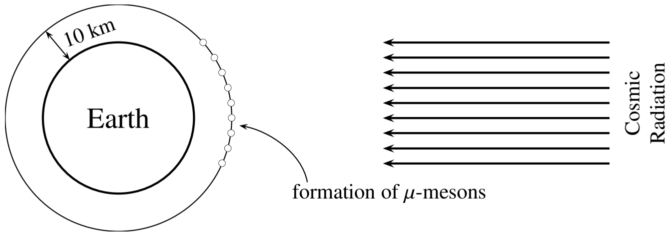


Figure 27.. μ-mesons created by cosmic radiation

What is the explanation for this?
Explanation from our point of view on earth: Time dilation implies that time goes by much slower for the muons moving with very high speed towards the earth. For this reason, from our point of view, the lifetime of muons is much longer than $2 \cdot 10^{-6}$ s.
Explanation from the muon's point of view: Because of length contraction the distance between the outer atmosphere and the surface of the earth is much less than 10 km. Therefore the distance to the surface can be overcome even in the short time at disposal.
This example shows nicely that length contraction and time dilation are really two sides of the same medal.

We now consider an observer B that is subject to acceleration. We assume that B always has velocity below light speed with respect to inertial observers, i.e. its world line is timelike. Parametrize the world line of B by $\mathbf{x} : [a, b] \to \mathbb{R}^4$. After possibly using the parameter transform $s \mapsto -s$, we can assume that $\mathbf{x}$ is future directed, i.e. that $\frac{d\mathbf{x}}{ds}(s) \in \mathcal{Z}^\uparrow$ for all $s \in [a, b]$.

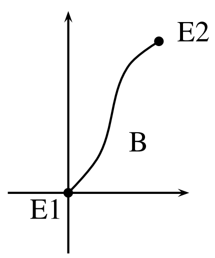


Figure 28.. Worldline of accelerated particle

What is the time elapsed between two events E1= $\mathbf{x}(a)$ and E2= $\mathbf{x}(b)$, measured on a clock taken along by observer B? In the special case that B moves with constant velocity (with respect to inertial observers), we already know that the time elapsed between E1 and E2 is given by

$$\sqrt{-\langle\!\langle E2-E1, E2-E1\rangle\!\rangle}.$$

We reduce the general case to this one. For a sufficiently fine partition $a = s_0 < s_1 < \ldots < s_n = b$ we have $\mathbf{x}(s_i) - \mathbf{x}(s_{i-i}) \in \mathcal{Z}^{\uparrow}$, $i = 1, \ldots, n$, because

$$\frac{\mathbf{x}(s_i) - \mathbf{x}(s_{i-i})}{s_i - s_{i-i}} \to \underbrace{\frac{d\mathbf{x}}{ds}}_{\in \mathcal{Z}^{\uparrow}},$$

as the mesh of the partition tends to 0. Since $\mathcal{Z}^{\uparrow}$ is open, $\frac{\mathbf{x}(s_i)-\mathbf{x}(s_{i-i})}{s_i - s_{i-i}}$ has to be in $\mathcal{Z}^{\uparrow}$ if $s_i - s_{i-1}$ is small enough[1].

This partition leads to the approximation of an accelerated observer B by a “piecewiese inertial observer”. This approximation becomes better as the mesh of the partition gets smaller. The time elapsed between to subsequent events $\mathbf{x}(s_{i-1})$ and $\mathbf{x}(s_i)$ in the view of the corresponding inertial observer with world line $\mathbb{R} \cdot (\mathbf{x}(s_i) - \mathbf{x}(s_{i-1})) + \mathbf{x}(s_{i-1})$ is

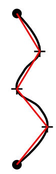


Figure 29.. Approximation by polygon

$$\sqrt{-\langle\!\langle \mathbf{x}(s_i) - \mathbf{x}(s_{i-1}), \mathbf{x}(s_i) - \mathbf{x}(s_{i-1})\rangle\!\rangle}.$$

Summation gives an approximate value for the time elapsed between E1= $\mathbf{x}(a)$ and E2= $\mathbf{x}(b)$ from the viewpoint of B:

$$\begin{aligned}
&\sum_{i=1}^{n} \sqrt{-\langle\!\langle \mathbf{x}(s_i) - \mathbf{x}(s_{i-1}), \mathbf{x}(s_i) - \mathbf{x}(s_{i-1})\rangle\!\rangle} \\
&= \sum_{i=1}^{n} \sqrt{-\left\langle\!\!\left\langle \frac{\mathbf{x}(s_i) - \mathbf{x}(s_{i-1})}{s_i - s_{i-1}}, \frac{\mathbf{x}(s_i) - \mathbf{x}(s_{i-1})}{s_i - s_{i-1}} \right\rangle\!\!\right\rangle} \cdot (s_i - s_{i-1}) \\
&\to \int_a^b \sqrt{-\left\langle\!\!\left\langle \frac{d\mathbf{x}}{ds}(s), \frac{d\mathbf{x}}{ds}(s) \right\rangle\!\!\right\rangle}\, ds,
\end{aligned}$$

as the mesh tends to 0 by the theorem on Riemann sums. We summarize: From the view of an accelerated observer B with word line $\mathbf{x}$, the time elapsed between the events $\mathbf{x}(a)$ and $\mathbf{x}(b)$ is given by

$$\boxed{\int_a^b \sqrt{-\left\langle\!\!\left\langle \frac{d\mathbf{x}}{ds}(s), \frac{d\mathbf{x}}{ds}(s) \right\rangle\!\!\right\rangle}\, ds}$$

[1] In fact, it is not hard to see that for every future-directed timelike curve $\mathbf{x}$, we always have $\mathbf{x}(t_2) - \mathbf{x}(t_1) \in \mathcal{Z}^{\uparrow}$ even if t_2 is much larger than t_1. However, we will not need this.

Definition 1.23. A future-directed parametrization $\mathbf{x} : [a, b] \to \mathbb{R}^4$ of a timelike world line is called a *parametrization by proper time* if

$$\left\langle \frac{d\mathbf{x}}{d\tau}, \frac{d\mathbf{x}}{d\tau} \right\rangle \equiv -1.$$

In other words, we have for all τ:

$$\frac{d\mathbf{x}}{d\tau}(\tau) \in H^3.$$

Remark 1.24. If the world line of an observer is parametrized by proper time, the parameter τ always gives the time elapsed from the view of this observer:

$$\int_a^{\tau_0} \sqrt{-\left\langle \frac{d\mathbf{x}}{d\tau}, \frac{d\mathbf{x}}{d\tau} \right\rangle} d\tau = \tau_0 - a.$$

Lemma 1.25. *Every timelike world line can be parametrized by proper time. The parametrization by proper time is unique up to parameter transformations of the form $\tau \mapsto \tau + \tau_0$ for fixed $\tau_0 \in \mathbb{R}$.*

Proof. *Existence:* Let $s \mapsto \mathbf{x}(s)$ a parametrization of the world line. Without loss of generality let $\frac{dx^0}{ds} > 0$, otherwise replace s by $-s$. For fixed $t_0 \in \mathbb{R}$ set

$$\psi(s) := \int_{t_0}^{s} \sqrt{-\left\langle \frac{d\mathbf{x}}{ds}(t), \frac{d\mathbf{x}}{ds}(t) \right\rangle} dt.$$

Then

$$\psi'(s) = \sqrt{-\left\langle \frac{d\mathbf{x}}{ds}(s), \frac{d\mathbf{x}}{ds}(s) \right\rangle} > 0.$$

Hence ψ is strictly monotonically increasing. For the inverse $\varphi := \psi^{-1}$ we have

$$\frac{d\varphi}{d\tau}(\tau) = \frac{1}{\psi'(\varphi(\tau))} = \frac{1}{\sqrt{-\left\langle \frac{d\mathbf{x}}{ds}(\varphi(\tau)), \frac{d\mathbf{x}}{ds}(\varphi(\tau)) \right\rangle}}.$$

This implies

$$\begin{aligned}
\left\langle \frac{d(\mathbf{x} \circ \varphi)}{d\tau}(\tau), \frac{d(\mathbf{x} \circ \varphi)}{d\tau}(\tau) \right\rangle &= \left\langle \frac{d\mathbf{x}}{ds}(\varphi(\tau)) \cdot \frac{d\varphi}{d\tau}(\tau), \frac{d\mathbf{x}}{ds}(\varphi(\tau)) \cdot \frac{d\varphi}{d\tau}(\tau) \right\rangle \\
&= \left(\frac{d\varphi}{d\tau}(\tau) \right)^2 \cdot \left\langle \frac{d\mathbf{x}}{ds}(\varphi(\tau)), \frac{d\mathbf{x}}{ds}(\varphi(\tau)) \right\rangle \\
&= -1.
\end{aligned}$$

Uniqueness: Let $\mathbf{x}$ and $\mathbf{x} \circ \varphi$ be parametrizations by proper time. Then

$$-1 = \left\langle \frac{d(\mathbf{x} \circ \varphi)}{d\tau}, \frac{d(\mathbf{x} \circ \varphi)}{d\tau} \right\rangle = \left(\frac{d\varphi}{d\tau}\right)^2 \cdot \left\langle \frac{d\mathbf{x}}{ds}, \frac{d\mathbf{x}}{ds} \right\rangle = -\left(\frac{d\varphi}{d\tau}\right)^2 .$$

This implies $\left|\frac{d\varphi}{d\tau}\right| = 1$ and hence $\varphi(\tau) = \pm\tau + \tau_0$ for some fixed $\tau_0 \in \mathbb{R}$. Since both parametrizations are future directed, we have

$$0 < \frac{d(\mathbf{x} \circ \varphi)^0}{d\tau} = \frac{d\varphi}{d\tau} \cdot \underbrace{\frac{dx^0}{ds}}_{>0} \quad \text{and hence} \quad \frac{d\varphi}{d\tau} > 0.$$

Thus $\varphi(\tau) = \tau + \tau_0$. □

The Twin Paradox

Suppose Alice and Bob are twins. Bob decides to go on a round trip in a space craft while Alice remains at rest at home in an inertial frame. On his return, Bob is younger than Alice!

This can be seen as follows: In the inertial frame of Alice, let E1= $\mathbf{0}$ the event of Bob's departure and E2= $(T, 0, 0, 0)^\top$ the event of his return. This means that Alice has aged by time T during the separation of the twins.

To compute the aging of Bob, let $s \mapsto \mathbf{x}(s) = (s, \hat{\mathbf{x}}(s))$ be a parametrization of Bob's worldline in the inertial system of Alice. We compute the time that has passed for Bob:

$$\int_0^T \sqrt{-\left\langle \frac{d\mathbf{x}}{ds}, \frac{d\mathbf{x}}{ds} \right\rangle} ds = \int_0^T \underbrace{\sqrt{1 - \left\| \frac{d\hat{\mathbf{x}}}{ds} \right\|^2}}_{\substack{\leq 1 \text{ and } =1 \text{ only} \\ \text{if } \| \frac{d\hat{\mathbf{x}}}{ds} \| = 0}} ds < T.$$

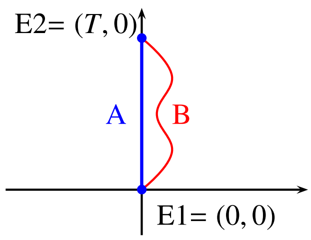


Figure 30.. Twin paradox

We conclude: Travel keeps you young!

This phenomenon was verified experimentally. In the *Hafele-Keating experiment (1971)* [4, 5], four atomic clocks were flown on commerical flights around the world, once eastward and once westward. A clock in an inertial system of the center of the earth would travel eastward in the direction of the rotation of the earth. In the experiment all clocks are travelling but when travelling eastward they are closer to an inertial observer than when travelling westward. From the flight paths of the trips, our relativistic kinematic formulas predict that the westward travelling clocks should have gained 280 nonseconds compared to the eastward trip. The measurements gave 332 nanoseconds.

The difference is mostly due to another relativistic effect, namely the influence of gravitation. Gravitation is weaker on board of the airplane while in high altitude and this has an impact on time as we will see.

Definition 1.26. Let $\mathbf{x} : I \to \mathbb{R}^4$ by a parametrization by proper time of the world line of a timelike particle. The vector

$$\mathbf{u} := \frac{d\mathbf{x}}{d\tau}$$

is called *four-velocity* of the particle (at $\mathbf{x}(\tau)$) and

$$\mathbf{a} := \frac{d^2\mathbf{x}}{d\tau^2}$$

is called its *four-acceleration*.

Remark 1.27. By definition of a proper-time parametrization, the four-velocity is a curve in H^3. Hence its derivative, the four-acceleration, is always tangent to H^3,

$$\mathbf{a}(\tau) = \frac{d\mathbf{u}}{d\tau}(\tau) \in T_{\mathbf{u}(\tau)}H^3 = \mathbf{u}(\tau)^{\perp\!\!\!\perp}.$$

In particular, by Lemma 1.12, four-acceleration is always spacelike.

Write $\mathbf{x} = (x^0, \hat{\mathbf{x}})$ for the world line of a particle. The *observed velocity* of $\mathbf{x}$ from the view of an inertial observer with world line $\mathbb{R}\mathbf{e_0}$ is given by

$$\frac{\hat{\mathbf{u}}}{u^0},$$

as discussed before. The *observed acceleration* from the viewpoint of this inertial observer is the change of velocity per change of time, which is

$$\frac{d}{dx^0}\left(\frac{\hat{\mathbf{u}}}{u^0}\right) = \frac{1}{u^0}\frac{d}{d\tau}\left(\frac{\hat{\mathbf{u}}}{u^0}\right) = \frac{1}{u^0}\frac{\frac{d}{d\tau}\hat{\mathbf{u}} \cdot u^0 - \frac{d}{d\tau}u^0 \cdot \hat{\mathbf{u}}}{(u^0)^2} = \frac{\hat{\mathbf{a}}}{(u^0)^2} - \frac{a^0}{(u^0)^3}\hat{\mathbf{u}}.$$

If the inertial frame is the rest frame of the particle at time $\tau = \tau_0$, i.e., $\mathbf{u}(\tau_0) = \mathbf{e_0}$, then the four-acceleration satisfies $\mathbf{a}(\tau_0) = (0, \hat{\mathbf{a}}(\tau_0))$ because $\mathbf{a}(\tau_0) \perp\!\!\!\perp \mathbf{u}(\tau_0) = \mathbf{e_0}$ and hence $a^0(\tau_0) = 0$. Since $u^0(\tau_0) = 1$ and $a^0(\tau_0) = 0$, the observed acceleration is just $\hat{\mathbf{a}}(\tau_0)$. The absolute value of the observed acceleration in the rest frame is therefore

$$|\hat{\mathbf{a}}(\tau_0)| = \sqrt{\langle\!\langle \mathbf{a}(\tau_0), \mathbf{a}(\tau_0)\rangle\!\rangle}.$$

1.5. Mass and energy

Definition 1.28. A *force field* $\mathbf{F}$ is a smooth mapping $\mathbf{F} : \mathbb{R}^4 \times H^3 \to \mathbb{R}^4$ such that for all $\mathbf{x} \in \mathbb{R}^4$ and $\mathbf{u} \in H^3$ we have

$$\langle\!\langle \mathbf{F}(\mathbf{x}, \mathbf{u}), \mathbf{u}\rangle\!\rangle = 0.$$

This means $\mathbf{F}(\mathbf{x}, \mathbf{u}) \in T_{\mathbf{u}}H^3$. We impose this condition because we know already that it is fulfilled for the four-acceleration and we want to demand later that the force is proportional to acceleration, as in Newton's second law.
Let $m_0 > 0$ be a constant which we interpret as the *rest mass* of a particle. Let the particle have world line $\mathbf{x}$ in an inertial system. The analog to the Newton's second law is then:

If the world line of a particle is parametrized by proper time and the particle is subject to the force $\mathbf{F}$*, then*

$$\frac{d}{d\tau}(m_0 \mathbf{u}(\tau)) = \mathbf{F}(\mathbf{x}(\tau), \mathbf{u}(\tau)), \tag{1.8}$$

or, equivalently,

$$m_0 \frac{d^2}{d\tau^2}\mathbf{x}(\tau) = \mathbf{F}\Big(\mathbf{x}(\tau), \frac{d\mathbf{x}}{d\tau}(\tau)\Big).$$

This is an ordinary differential equation of second order for $\mathbf{x}$ if $\mathbf{F}$ is given. Given any initial conditions $\mathbf{x}(\tau_0)$ and $\mathbf{u}(\tau_0) = \frac{d\mathbf{x}}{d\tau}(\tau)$, it has a unique solution. Hence special relativity is, as the theory of classical mechanics, a deterministic theory.
In the rest frame of the particle, i.e. if $\mathbf{u}(\tau_0) = \mathbf{e_0}$, the relation $\mathbf{F}(\mathbf{x}(\tau_0), \mathbf{u}(\tau_0)) \perp\!\!\!\perp \mathbf{u}(\tau_\mathbf{0})$ means

$$\mathbf{F}(\mathbf{x}(\tau_0), \mathbf{u}(\tau_0)) = (0, \hat{\mathbf{F}}(\mathbf{x}(\tau_0), \mathbf{u}(\tau_0))).$$

Hence in the rest frame, we are left with the classical Newtonian equation of motion $m_0\, \hat{\mathbf{a}}(\tau_0) = \hat{\mathbf{F}}(\mathbf{x}(\tau_0), \mathbf{u}(\tau_0))$.
Without the assumption that the given inertial frame is the rest system of the particle, we define the *relativistic mass*

$$m(\tau) := \frac{m_0}{\sqrt{1 - \left|\frac{\hat{\mathbf{u}}(\tau)}{u^0(\tau)}\right|^2}}.$$

The inertial frame is the rest frame of the particle at the event $\mathbf{x}(\tau_0)$ if and only if $\mathbf{u}(\tau_0) = \mathbf{e_0}$, i.e. if and only if $m(\tau_0) = m_0$. Otherwise, we have $m(\tau) > m_0$. We then find

$$\frac{d}{dx^0}\left(m\frac{\hat{\mathbf{u}}}{u^0}\right) = m_0 \frac{d}{dx^0}\frac{\hat{\mathbf{u}}}{\sqrt{(u^0)^2 - |\hat{\mathbf{u}}|^2}} = m_0 \frac{d}{dx^0}\hat{\mathbf{u}} = \frac{m_0}{u^0}\frac{d}{d\tau}\hat{\mathbf{u}} = \frac{m_0}{u^0}\hat{\mathbf{a}} = \frac{\hat{\mathbf{F}}}{u^0}.$$

This is the classical Newtonian equation of motion with mass m and force $\frac{\hat{\mathbf{F}}}{u^0}$. Therefore, the relativistic mass is interpreted as the mass of the particle from the viewpoint of our inertial observer and $\frac{\hat{\mathbf{F}}}{u^0}$ as the *observed force* acting on the particle from the viewpoint of this observer.

Definition 1.29. For a particle with constant rest mass m_0 and a world line parametrized by proper time, the *four-momentum* is given by

$$\mathbf{p} := m_0 \cdot \mathbf{u}$$

where $\mathbf{u}$ denotes its four-velocity.

Equation (1.8) then takes the form

$$\frac{d}{d\tau}\mathbf{p}(\tau) = \mathbf{F}(\mathbf{x}(\tau), \mathbf{u}(\tau)).$$

As we have seen,

$$\frac{d}{dx^0}\left(\underset{\text{relativistic mass}}{m} \cdot \underset{\text{observed velocity}}{\frac{\hat{\mathbf{u}}}{u^0}}\right) = \underset{\text{observed force}}{\frac{\hat{\mathbf{F}}}{u^0}}$$

where $m = \frac{m_0}{\sqrt{1-\|\hat{\mathbf{u}}/u^0\|^2}}$. From

$$-1 = \langle\langle \mathbf{u}, \mathbf{u} \rangle\rangle = -(u^0)^2 + |\hat{\mathbf{u}}|^2$$

we find

$$\frac{1}{(u^0)^2} = 1 - \left|\frac{\hat{\mathbf{u}}}{u^0}\right|^2$$

and therefore

$$u^0 = \frac{1}{\sqrt{1 - \left|\frac{\hat{\mathbf{u}}}{u^0}\right|^2}}.$$

Hence we can write $m = u^0 m_0$ for the relativistic mass. The time component of the vector equation (1.8) is then

$$u^0 \frac{d}{dx^0} m = \frac{d}{d\tau}(m_0 \cdot u^0) \overset{(1.8)}{=} F^0(\mathbf{x}, \mathbf{u}) = \frac{1}{u^0}\left\langle \hat{\mathbf{F}}, \hat{\mathbf{u}} \right\rangle,$$

because $0 = \langle\langle \mathbf{F}(\mathbf{x}, \mathbf{u}), \mathbf{u} \rangle\rangle = -F^0 u^0 + \left\langle \hat{\mathbf{F}}, \hat{\mathbf{u}} \right\rangle$. This implies

$$\frac{d}{dx^0} m = \left\langle \frac{\hat{\mathbf{F}}}{u^0}, \frac{\hat{\mathbf{u}}}{u^0} \right\rangle.$$

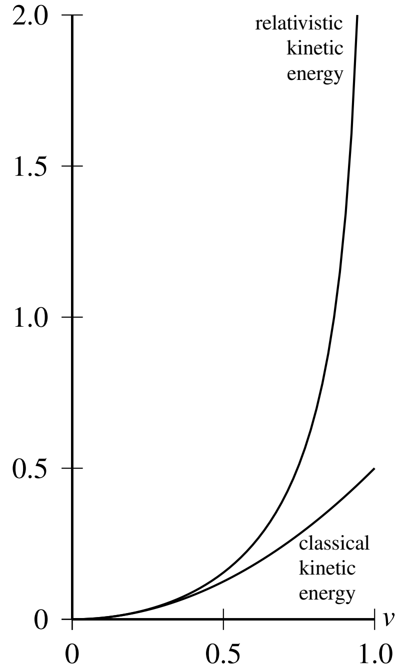


Figure 31.. Classical versus relativistic kinetic energy

This is the classical energy equation (1.2) with the relativistic mass m instead of the kinetic energy E. Therefore we can interpret the relativistic mass as the energy of the particle as well.

$$E = m = \frac{m_0}{\sqrt{1 - \left|\frac{\hat{\mathbf{u}}}{u^0}\right|^2}} = \underbrace{m_0}_{\substack{\text{rest}\\\text{energy}}} + \underbrace{\underbrace{\frac{m_0}{2}\left|\frac{\hat{\mathbf{u}}}{u^0}\right|^2}_{\substack{\text{classical}\\\text{kinetic}\\\text{energy}}} + \frac{3m_0}{8}\left|\frac{\hat{\mathbf{u}}}{u^0}\right|^4 + \mathrm{O}\left(\left|\frac{\hat{\mathbf{u}}}{u^0}\right|^6\right)}_{\text{relativistic kinetic energy}},$$

where we used that $\frac{1}{\sqrt{1-x}} = 1 + \frac{1}{2}x + \frac{3}{8}x^2 + \mathrm{O}(x^3)$. Since energy has has the same physical units as mass · velocity^2 we arrived at the famous formula

$$\boxed{E_{\text{phys}} = m_{\text{phys}} \cdot c^2}$$

So far, we only discussed point particles. In classical continuum mechanics extended bodies possess

(1) a *mass density*: $\varrho : \mathbb{R} \times \mathbb{R}^3 \to \mathbb{R}$; the total mass of a body at time t is then given by

$$\int_{\mathbb{R}^3} \varrho(t, \hat{\mathbf{x}})\, dx^1\, dx^2\, dx^3,$$

(2) a *momentum density*: $\mathbf{p} : \mathbb{R} \times \mathbb{R}^3 \to \mathbb{R}^3$; the total momentum of the body at time t is then given by

$$\int_{\mathbb{R}^3} \mathbf{p}(t, \hat{\mathbf{x}})\, dx^1\, dx^2\, dx^3,$$

(3) a *stress tensor*: $\sigma : \mathbb{R} \times \mathbb{R}^3 \to \{\text{symmetric bilinear forms on } \mathbb{R}^3\}$. To give a physical interpretation of σ let $\mathbf{b_1}, \mathbf{b_2}, \mathbf{b_3}$ be a diagonalizing orthonormal basis of $\sigma|_{(t,\hat{\mathbf{x}})}$. Then

$$\sigma|_{(t,\hat{\mathbf{x}})}(\mathbf{b_i}, \mathbf{b_j}) = \begin{cases} \lambda_i & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

At time t and at the point $\hat{\mathbf{x}}$, the body exerts pressure of strength λ_i in direction $\mathbf{b_i}$.

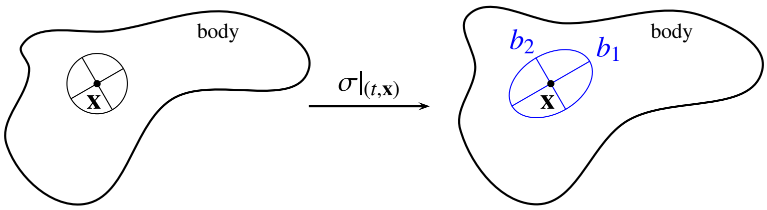


Figure 32.. Stress tensor

In relativity, these entities are combined to the *stress-energy tensor*

$$\mathbf{T} : \mathbb{R}^4 \to \{\text{symmetric bilinear forms on } \mathbb{R}^4\}, \quad \mathbf{T} = \left(\begin{array}{c|c} \varrho & \mathbf{p}^\top \\ \hline \mathbf{p} & \sigma \end{array}\right).$$

To every extended body we assign such a $\mathbf{T}$, with the following physical interpretation. For an observer B whose world line has the four-velocity $\mathbf{u}$ at the event $\mathbf{x}$,

$$\begin{aligned} \mathbf{T}|_{\mathbf{x}}(\mathbf{u}, \mathbf{u}) &= \text{mass density of the body at the event } \mathbf{x}\text{, as observed by B} \\ &= \text{energy density of the body at the event } \mathbf{x}\text{, as observed by B,} \\ \mathbf{T}|_{\mathbf{x}}(\mathbf{u}, \mathbf{e}) &= \langle\!\langle \text{momentum density of the body as observed by B at event } \mathbf{x}, \mathbf{e} \rangle\!\rangle \text{ where } \mathbf{e} \perp\!\!\!\perp \mathbf{u}, \\ \mathbf{T}|_{\mathbf{x}}(\mathbf{e}, \mathbf{e}') &= (\text{stress tensor of the body at event } \mathbf{x}\text{, as observed by B})(\mathbf{e}, \mathbf{e}') \text{ where } \mathbf{e}, \mathbf{e}' \perp\!\!\!\perp \mathbf{u}. \end{aligned}$$

Example 1.30. **(1)** *Vacuum:* $\mathbf{T} = 0$.
(2) *Dust:* Our universe is filled with dust. So we have a nonnegative mass density $\varrho : \mathbb{R}^4 \to \mathbb{R}$. The dust particles do not experience any pressure, hence no stresses. The four-velocities of the dust particles define a timelike, future-directed unit vector field $\mathbf{u}$.

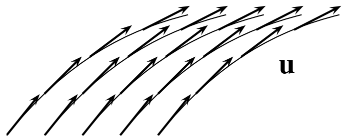


Figure 33.. Observer field

Since $\mathbf{p}$ is proportional to $\varrho\mathbf{u}$, an observer riding on a dust particle (i.e. with four-velocity $\mathbf{u}$) will observe $\mathbf{T}(\mathbf{u}, \mathbf{e})$ for any $\mathbf{e} \perp\!\!\!\perp \mathbf{u}$. Thus for observers with four-velocity $\mathbf{u}$, the momentum density and stresses vanishes. Hence

$$\mathbf{T}|_{\mathbf{x}}(\mathbf{y}, \mathbf{z}) = \varrho(\mathbf{x}) \langle\!\langle \mathbf{y}, \mathbf{u}(\mathbf{x}) \rangle\!\rangle \cdot \langle\!\langle \mathbf{z}, \mathbf{u}(\mathbf{x}) \rangle\!\rangle .$$

(3) *Ideal Liquid:* For an ideal liquid we also have a nonnegative mass density $\varrho : \mathbb{R}^4 \to \mathbb{R}$ and the worldline of the molecules yield an four-velocity vector field $\mathbf{u} : \mathbb{R}^4 \to H^3$. Again, the momentum density vanishes for an observer with four-velocity $\mathbf{u}$. In addition, we experience pressure which is isotropic, i.e. the same from all directions. Thus $\sigma|_{\mathbf{x}} = \lambda(\mathbf{x}) \langle \cdot, \cdot \rangle$. Again, the four-velocities of the liquid molecules define a timelike, future-directed unit vector field $\mathbf{u}$. Hence

$$\mathbf{T}|_{\mathbf{x}}(\mathbf{y}, \mathbf{z}) = (\varrho(\mathbf{x}) + \lambda(\mathbf{x})) \langle\!\langle \mathbf{y}, \mathbf{u}(\mathbf{x}) \rangle\!\rangle \cdot \langle\!\langle \mathbf{z}, \mathbf{u}(\mathbf{x}) \rangle\!\rangle + \lambda(\mathbf{x}) \langle\!\langle \mathbf{y}, \mathbf{z} \rangle\!\rangle .$$

Indeed, for $\mathbf{e}, \mathbf{e}' \in \mathbf{u}^{\perp\!\!\!\perp}$ we find

$$\begin{aligned}
\mathbf{T}|_{\mathbf{x}}(\mathbf{u}(\mathbf{x}), \mathbf{u}(\mathbf{x})) &= (\varrho(\mathbf{x}) + \lambda(\mathbf{x})) \langle\!\langle \mathbf{u}(\mathbf{x}), \mathbf{u}(\mathbf{x}) \rangle\!\rangle \cdot \langle\!\langle \mathbf{u}(\mathbf{x}), \mathbf{u}(\mathbf{x}) \rangle\!\rangle + \lambda(\mathbf{x}) \langle\!\langle \mathbf{u}(\mathbf{x}), \mathbf{u}(\mathbf{x}) \rangle\!\rangle \\
&= \varrho(\mathbf{x}) + \lambda(\mathbf{x}) - \lambda(\mathbf{x}) = \varrho(\mathbf{x}), \\
\mathbf{T}|_{\mathbf{x}}(\mathbf{u}(\mathbf{x}), \mathbf{e}) &= 0, \\
\mathbf{T}|_{\mathbf{x}}(\mathbf{e}, \mathbf{e}') &= \lambda(\mathbf{x}) \langle\!\langle \mathbf{e}, \mathbf{e}' \rangle\!\rangle .
\end{aligned}$$

Later we will learn about a more conceptual way of finding the energy stress tensor for different kinds of matter.

1.6. Closing remarks about special relativity

Let us summarize briefly the structure of special relativity, now making use of differential geometric language. Space and time are joined to the 4-dimensional spacetime. The Postulate of Special Relativity states that the mathematical model for spacetime is a time-oriented Lorentz manifold M which is isometric to $(\mathbb{R}^4, g_{\mathrm{Mink}})$. An isometry $M \to (\mathbb{R}^4, g_{\mathrm{Mink}})$ preserving the time orientation is called an inertial frame. The coordinates $(x^0, \hat{\mathbf{x}})$ that are assigned to an event by such an isometry are the time and space coordinates from the point of view of an observer with world line $\mathbb{R} \cdot \mathbf{e_0}$.

The world lines of particles slower than light are the timelike smooth curves in M. The world lines of particles moving at the speed of light are null curves. The world lines of particles not subject to any acceleration are geodesics (straight lines) in M.
Let $\mathcal{H} := \{\xi \in TM \mid g(\xi, \xi) = -1 \text{ and } \xi \text{ is future directed}\}$. An external force is given by a vector field $\mathbf{F}$ along the footpoint map $\pi : \mathcal{H} \to M$ with $g(\mathbf{F}(\xi), \xi) = 0$ for all $\xi \in \mathcal{H}$. We have an analog to Newton's second law,

$$m_0 \frac{\nabla}{d\tau}\frac{d\mathbf{x}}{d\tau} = m_0 \frac{d^2}{d\tau^2}\mathbf{x}(\tau) = \frac{d}{d\tau}(m_0\mathbf{u})(\tau) = F(\mathbf{x}(\tau), \mathbf{u}(\tau)).$$

This equation can be studied in arbitrary coordinate systems, not only in inertial frames.
All relevant physical objects possess a stress-energy tensor containing information about the mass density, momentum density and stress density.

1.7. Exercises

1.1. A spacecraft travels from earth to a distant object X, its rear engine inducing constant acceleration (=gravitational acceleration) $g = 9.81ms^{-2}$. At half the distance, the spacecraft is turned over (so its rear engine now induces the same deceleration).
According to classical kinematics, how long does the journey take and what is the maximal velocity for
(a) X = moon (400.000 km),
(b) X = mars (56-400 million km),
(c) X = Proxima Centauri (4,3 light years) und
(d) X = Andromeda galaxy (2 million lightyears)?
Remember: 1 light year $\approx 9.461 \cdot 10^{12} km$.

1.2. Show that for $\mathbf{L} \in \mathrm{Mat}(4 \times 4, \mathbb{R})$ the following are equivalent:

$$\mathbf{L} \cdot \mathbf{I}_{1,3} \cdot \mathbf{L}^\top = \mathbf{I}_{1,3} \quad \Leftrightarrow \quad \mathbf{L}^\top \cdot \mathbf{I}_{1,3} \cdot \mathbf{L} = \mathbf{I}_{1,3}.$$

1.3. (a) Show that each Galilean transformation is invertible and that its inverse is again a Galilean transformation.
(b) Show that the set of Galilean tranformations with composition forms a group.
(c) Show that each Lorentz transformation is invertible and that its inverse is again a Lorentz transformation.
(d) Show that $\mathcal{L}$ with composition forms a group.
(e) Show that each Poincaré transformation is invertible and that its inverse is again a Poincaré transformation.
(f) Show that $\mathcal{P}$ with composition forms a group.

1.4. Determine all Poincaré transformations of $\mathbb{R}^4$ which are also Galilean transformations. Why is this subgroup of transformations not sufficient to derive (1.1)?

1.5. (a) Show that $\mathcal{L}_+^\uparrow \cdot \mathcal{L}_+^\uparrow \subset \mathcal{L}_+^\uparrow$, $\mathcal{L}_+^\uparrow \cdot \mathcal{L}_+^\downarrow \subset \mathcal{L}_+^\downarrow$, and similarly for all other combinations.
(b) Conclude from a) that $\mathcal{L}_+^\uparrow$, $\mathcal{L}_+$, $\mathcal{L}^\uparrow$, and $\mathcal{L}_+^\uparrow \sqcup \mathcal{L}_-^\downarrow$ are subgroups of $\mathcal{L}$.

1.6. (a) Let $\mathbf{L} \in \mathcal{L}$ with $\mathbf{L}\mathbf{e_0} = \mathbf{e_0}$. Show that $\mathbf{L}$ is of the form

$$\mathbf{L} = \left(\begin{array}{c|c} 1 & 0 \\ \hline 0 & \mathbf{A} \end{array}\right)$$

with $\mathbf{A} \in \mathrm{O}(3)$.
(b) Let $\mathbf{L} \in \mathcal{L}_+^\uparrow$ with $\mathbf{L}\mathbf{e_2} = \mathbf{e_2}$ and $\mathbf{L}\mathbf{e_3} = \mathbf{e_3}$. Show that $\mathbf{L}$ is a boost.

1.7. Use the law of cosines for sides to show that the sum of angles in a hyperbolic triangle is smaller than 180°,

$$\alpha + \beta + \gamma < \pi.$$

1.8. Let B1 and B2 be two inertial observers. Let P be a Poincaré transformation with linear part L (a Lorentz transformation). Suppose B2 has the world line $P(\mathbb{R}\mathbf{e_0})$ in the inertial system of B1. Suppose furthermore that there is an external force with force field $\mathbf{F}$ from the viewpoint of B1 and $\tilde{\mathbf{F}}$ from the viewpoint of B2.

(a) Prove the relation

$$\tilde{\mathbf{F}}(\tilde{\mathbf{x}}, \tilde{\mathbf{u}}) = L^{-1} \cdot \mathbf{F}(P\tilde{\mathbf{x}}, L \cdot \tilde{\mathbf{u}}).$$

(b) For an electromagnetic field show the corresponding relation

$$\tilde{\mathcal{F}}|_{\tilde{\mathbf{x}}}(\tilde{\mathbf{u}}, \tilde{\mathbf{v}}) = \mathcal{F}|_{P\tilde{\mathbf{x}}}(L\tilde{\mathbf{u}}, L\tilde{\mathbf{v}}).$$

1.9. Inertial observer B observes the constant electric field $\hat{\mathbf{E}} = (1, 0, 0)$ but no magnetic field, $\hat{\mathbf{B}} = (0, 0, 0)$.
(a) In B's inertial system, compute the world line $\mathbf{x}$ of a particle with charge q, rest mass m_0, $\mathbf{x}(0) = (0, 0, 0, 0)$ and $\frac{d\mathbf{x}}{d\tau}(0) = \mathbf{e_0}$.
(b) Is there an inertial observer B2 who will observe a vanishing electric field $\tilde{\hat{\mathbf{E}}} = (0, 0, 0)$ in this situation?

1.10. Let P be a planet being at rest in an inertial system. Planet P has a moon M circling around P at constant speed and constant distance (as observed by P). A spaceship S starts on P and moves with constant velocity with respect to P, hence defining a second inertial system. Do observes on S also observe a constant distance between P and M? If not, what kind of curve is the orbit of M around P as observed by S?

Discuss these questions
(a) when S moves perpendicularly to the plane E containing the orbit of the moon,
(b) when S moves within E.

1.11. Two trains are standing on the same track at a certain distance. They are connected by a tight rod. Both trains depart at the same time and start traveling at the same constant acceleration. Will the rod tear?

1.12. Albert is traveling at night in a train that is moving at constant velocity v. As usual there are problems when traveling with Deutsche Bahn; this time there is an electricity failure so that it becomes totally dark. Albert turns on his flash light. Near which seat does he have to position himself so that his light reaches the tip and the rear of the train at the same time, once from his point of view and once from the point of view of an observer watching the train from outside?

Figure 34.. Train travel

Answer these questions
(a) for $v = c/2$,
(b) for $v = 4c/5$.

1.13. Let $s \mapsto \mathbf{x}(s)$ be a smooth parametrized curve in $\mathbb{R}^4$.
(a) Show: if $\mathbf{x}$ is timelike, i.e., if $\dot{\mathbf{x}}(s)$ is timelike for all s, then $\mathbf{x}(s_2) - \mathbf{x}(s_1)$ is timelike whenever $s_1 \neq s_2$.
(b) Show: if, in addition, $\mathbf{x}$ is future directed, i.e., if $\dot{\mathbf{x}}(s) \in \mathcal{Z}^\uparrow$ for all s, then $\mathbf{x}(s_2) - \mathbf{x}(s_1)$ is future directed whenever $s_1 < s_2$.
(c) Does (a) also hold if "timelike" is replaced by "lightlike"?

1.14. Let m_0 be the rest mass and m the relativistic mass of a particle w.r.t. an inertial observer. Show that the observed momentum $\hat{\mathbf{p}}$ of the particle satisfies

$$\|\hat{\mathbf{p}}\| = \sqrt{m^2 - m_0^2}.$$

1.15. Redo Exercise 1.1, now using relativity theory instead of Newtonian mechanics. Compute the travel times from the viewpoint of the crew on board the spacecraft as well as from the viewpoint of those left on earth. Moreover, compute the maximal velocities relative to the earth that the spacecraft reaches during the journey.[2]

[2]The cosmic travel planner at https://math.cbaer.eu/CTP/CTP.html can be used to compute these travel times and maximal velocities for arbitrary destinations.

2. Einstein's field equations

The goal is now to include gravitation into relativity theory. From now on the reader will be assumed to be familiar with differential geometry. We start by quickly reviewing classical Newtonian gravity theory.

2.1. Classical theory of gravitation

Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^3$ be the position vectors of two point particles with masses m and M, respectively. *Newton's law of gravitation* says that in a Galilean inertial frame $\mathbf{y}$ exerts the force

$$\mathbf{F} = -\frac{GmM}{|\mathbf{x}-\mathbf{y}|^2}\frac{\mathbf{x}-\mathbf{y}}{|\mathbf{x}-\mathbf{y}|} \tag{2.1}$$

on $\mathbf{x}$. Here $G = 6,673 \cdot 10^{-11}\ \mathrm{m}^3\mathrm{kg}^{-1}\mathrm{s}^{-2}$ is the *gravitational constant*.
We make the simplifying assumption that $\mathbf{y}$ is fixed at $\mathbf{y} = \mathbf{0} \in \mathbb{R}^3$. This can be justified for instance if $M \gg m$. Since the gravitational force exerted by $\mathbf{x}$ and $\mathbf{y}$ on one another has the same absolute value, it will have a strong impact on the light body $\mathbf{x}$ but hardly affect the heavy body $\mathbf{y}$. Combining the law of gravitation and Newton's second law $\mathbf{F} = m\ddot{\mathbf{x}}$ we get

$$\ddot{\mathbf{x}} = -\frac{GM}{|\mathbf{x}|^3}\mathbf{x}. \tag{2.2}$$

Remark 2.1. The mass m of $\mathbf{x}$ has canceled in (2.2), so the orbit of $\mathbf{x}$ does not depend on its mass. A priori, one would have to distinguish between the inertial mass m_{inert} occurring in Newton's second law $\mathbf{F} = m_{\mathrm{inert}} \cdot \ddot{\mathbf{x}}$ and the gravitational mass m_{grav} in

$$\mathbf{F} = -\frac{Gm_{\mathrm{grav}}M_{\mathrm{grav}}}{|\mathbf{x}-\mathbf{y}|^2}\frac{\mathbf{x}-\mathbf{y}}{|\mathbf{x}-\mathbf{y}|}.$$

Equation (2.2) and hence the equality $m_{\mathrm{inert}} = m_{\mathrm{grav}}$ of these two concepts of mass is experimentally well tested (see https://www.youtube.com/watch?v=5C5_dOEyAfk) and is therefore considered an empirical fact.

Define the *angular momentum per mass* by $\mathbf{L}(t) := \mathbf{x}(t) \times \dot{\mathbf{x}}(t)$.

Lemma 2.2 (Preservation of angular momentum). *If* $\mathbf{x}$ *satisfies equation* (2.2) *then* $\mathbf{L}$ *is constant.*

Proof. We compute

$$\frac{d}{dt}\mathbf{L} = \underbrace{\dot{\mathbf{x}} \times \dot{\mathbf{x}}}_{=\mathbf{0}} + \mathbf{x} \times \ddot{\mathbf{x}} \overset{(2.2)}{=} -\frac{GM}{|\mathbf{x}|^3} \underbrace{\mathbf{x} \times \mathbf{x}}_{=\mathbf{0}} = \mathbf{0}. \qquad \square$$

Remark 2.3. Assume that $\mathbf{x}$ satisfies (2.2) so that $\mathbf{L}$ is constant. If $\mathbf{L} \neq \mathbf{0}$, then $\mathbf{x}(t) \perp \mathbf{L}$ for all t. Hence $\mathbf{x}$ is confined to the plane perpendicular to $\mathbf{L}$.
If $\mathbf{L} = \mathbf{0}$, then $\mathbf{x}(t) = \mathbf{0}$ or $\dot{\mathbf{x}}(t) = \lambda(t)\mathbf{x}(t)$, that is

$$\mathbf{x}(t) = \mathbf{x}(t_0) \cdot e^{\int_{t_0}^{t} \lambda(s)\,ds}.$$

This means $\mathbf{x}(t)$ lies on the straight line through $\mathbf{0}$ and $\mathbf{x}(t_0)$ (with t_0 fixed). In this case $\mathbf{x}$ is even confined to a one-dimensional subspace.

Let $\mathbf{L} \neq \mathbf{0}$. We choose the coordinate system such that $\mathbf{L} = \|\mathbf{L}\| \cdot \mathbf{e_3}$. Hence $\mathbf{x}$ stays in the $\mathbf{e_1}$-$\mathbf{e_2}$-plane. We introduce polar coordinates (r, φ) in the $\mathbf{e_1}$-$\mathbf{e_2}$-plane:

$$x^1 = r\cos\varphi \quad \text{and} \quad x^2 = r\sin\varphi.$$

We express (2.2) in polar coordinates:

$$\begin{aligned}
\ddot{\mathbf{x}} = \frac{\nabla}{dt}\dot{\mathbf{x}} &= \frac{\nabla}{dt}\left(\dot{r}\frac{\partial}{\partial r} + \dot{\varphi}\frac{\partial}{\partial \varphi}\right) \\
\text{covariant derivative w.r.t. } g_{\text{eukl}} \quad &= \ddot{r}\frac{\partial}{\partial r} + \dot{r}\frac{\nabla}{dt}\frac{\partial}{\partial r} + \ddot{\varphi}\frac{\partial}{\partial \varphi} + \dot{\varphi}\frac{\nabla}{dt}\frac{\partial}{\partial \varphi} \\
&= \ddot{r}\frac{\partial}{\partial r} + \dot{r}\nabla_{\dot{r}\frac{\partial}{\partial r}+\dot{\varphi}\frac{\partial}{\partial \varphi}}\frac{\partial}{\partial r} + \ddot{\varphi}\frac{\partial}{\partial \varphi} + \dot{\varphi}\nabla_{\dot{r}\frac{\partial}{\partial r}+\dot{\varphi}\frac{\partial}{\partial \varphi}}\frac{\partial}{\partial \varphi} \qquad (2.3)
\end{aligned}$$

In polar coordinates (r, φ) the metric coefficients of the Euclidean metric g_{eukl} are

$$(g_{ij}) = \begin{pmatrix} 1 & 0 \\ 0 & r^2 \end{pmatrix}$$

and the Christoffel symbols are easily computed to be

$$\Gamma_{12}^2 = \Gamma_{21}^2 = \frac{1}{r}, \quad \Gamma_{22}^1 = -r, \text{ and all other } \Gamma_{ij}^k = 0.$$

Therefore

$$\nabla_{\dot{r}\frac{\partial}{\partial r}+\dot{\varphi}\frac{\partial}{\partial \varphi}}\frac{\partial}{\partial r} = \dot{r}\underbrace{\nabla_{\frac{\partial}{\partial r}}\frac{\partial}{\partial r}}_{=\mathbf{0}} + \dot{\varphi}\underbrace{\nabla_{\frac{\partial}{\partial \varphi}}\frac{\partial}{\partial r}}_{=\frac{1}{r}\frac{\partial}{\partial \varphi}} = \frac{\dot{\varphi}}{r}\frac{\partial}{\partial \varphi}$$

and

$$\nabla_{\dot r\frac{\partial}{\partial r}+\dot\varphi\frac{\partial}{\partial\varphi}}\frac{\partial}{\partial\varphi} = \dot r\underbrace{\nabla_{\frac{\partial}{\partial r}}\frac{\partial}{\partial\varphi}}_{=\frac{1}{r}\frac{\partial}{\partial\varphi}} + \dot\varphi\underbrace{\nabla_{\frac{\partial}{\partial\varphi}}\frac{\partial}{\partial\varphi}}_{=-r\frac{\partial}{\partial r}} = \frac{\dot r}{r}\frac{\partial}{\partial\varphi} - \dot\varphi r\frac{\partial}{\partial r}.$$

Inserting this into (2.3) yields

$$\frac{\nabla}{dt}\dot{\mathbf{x}} = (\ddot r - \dot\varphi^2 r)\frac{\partial}{\partial r} + \left(\ddot\varphi + 2\dot\varphi\frac{\dot r}{r}\right)\frac{\partial}{\partial\varphi}.$$

Now (2.2) reads

$$\frac{\nabla}{dt}\dot{\mathbf{x}} = -\frac{GM}{r^3}r\frac{\partial}{\partial r} = -\frac{GM}{r^2}\frac{\partial}{\partial r}.$$

so that (2.2) takes the form

$$\ddot r - \dot\varphi^2 r = -\frac{GM}{r^2}, \quad \ddot\varphi + 2\dot\varphi\frac{\dot r}{r} = 0 \tag{2.4}$$

in polar coordinates.

Lemma 2.4 (Kepler's Second Law). *Let* $\mathbf{x}$ *satisfy* (2.2)*. Then*

$$r^2\dot\varphi = \pm|\mathbf{L}|$$

is constant.

Proof. We compute

$$\begin{aligned}
|\mathbf{L}| &= |\mathbf{x}\times\dot{\mathbf{x}}| \\
&= \left|r\frac{\partial}{\partial r}\times\left(\dot r\frac{\partial}{\partial r}+\dot\varphi\frac{\partial}{\partial\varphi}\right)\right| \\
&= r|\dot\varphi|\cdot\left|\frac{\partial}{\partial r}\times\frac{\partial}{\partial\varphi}\right| \\
&= r|\dot\varphi|\cdot\underbrace{\left|\frac{\partial}{\partial r}\right|}_{=1}\cdot\underbrace{\left|\frac{\partial}{\partial\varphi}\right|}_{=r} \\
&= r^2|\dot\varphi|.
\end{aligned}$$

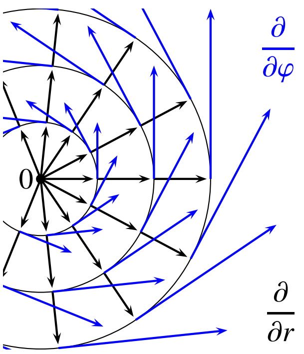


Figure 35.. Polar coordinate fields

□

After possibly applying a reflection, we can w.l.o.g. assume $r^2\dot\varphi = |\mathbf{L}|$.

Remark 2.5. Kepler's second law is often formulated in a more geometrical way as follows: The line segment from $\mathbf{0}$ to the point $\mathbf{x}(t)$ sweeps out equal areas during equal intervals of time.

To see this, we compute the area of the surface that is bordered by the line segments from $\mathbf{0}$ to $\mathbf{x}(t_0)$ and $\mathbf{x}(t_1)$ respectively ($t_0 < t_1$) and the corresponding segment of the orbit.
From differential geometry, it is known that the area element is given in polar coordinates by

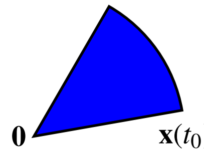


Figure 36.. Kepler's second law

$$\sqrt{\det(g_{ij})}\,dr\,d\varphi = \sqrt{\det\begin{pmatrix}1 & 0\\ 0 & r^2\end{pmatrix}}\,dr\,d\varphi = r\,dr\,d\varphi.$$

Employing the substitution rule for integration, we find for the area

$$\int_{\varphi(t_0)}^{\varphi(t_1)}\int_0^{r(\varphi)} r\,dr\,d\varphi = \int_{\varphi(t_0)}^{\varphi(t_1)} \frac{r(\varphi)^2}{2}\,d\varphi = \frac{1}{2}\int_{t_0}^{t_1} r(t)^2\,\dot\varphi(t)\,dt = \frac{|L|}{2}(t_1 - t_0).$$

So indeed, the area swept out is proportional to the time elapsed.

Now we restrict ourselves to the interesting case $\mathbf{L} \neq \mathbf{0}$. By Lemma 2.4 we have $r(t) > 0$ and $\dot\varphi(t) \neq 0$ for all t. Thus φ is strictly monotonic and we can define the auxiliary smooth function $u : I \to \mathbb{R}$, given by

$$u(s) := \frac{1}{r(\varphi^{-1}(s))} \qquad \text{i.e.,} \qquad u(\varphi(t)) = \frac{1}{r(t)}.$$

For the sake of brevity we write a dot for $\frac{d}{dt}$ and a prime for $\frac{d}{ds}$.

Lemma 2.6 (Orbit Equation). *Let* $\mathbf{x}$ *satisfy* (2.2). *Then we have*

$$u'' + u = \frac{GM}{|\mathbf{L}|^2}.$$

Proof. From $r^2\dot\varphi = |\mathbf{L}|$ we see

$$\dot\varphi(t) = \frac{|\mathbf{L}|}{r(t)^2} = |\mathbf{L}| \cdot u(\varphi(t))^2.$$

Hence

$$\dot r(t) = -\frac{u'(\varphi(t))}{u(\varphi(t))^2}\cdot\dot\varphi(t) = -\,|\mathbf{L}|\cdot u'(\varphi(t))$$

and therefore

$$\ddot{r}(t) = -|\mathbf{L}| \cdot u''(\varphi(t)) \cdot \dot{\varphi}(t) = -|\mathbf{L}|^2 \cdot u''(\varphi(t)) \cdot u(\varphi(t))^2.$$

Inserting this into (2.4) yields

$$-GMu^2 = -|\mathbf{L}|^2 u''u^2 - |\mathbf{L}|^2 u^4 \frac{1}{u} = -|\mathbf{L}|^2 u^2(u'' + u). \qquad \square$$

The orbit equation can be solved explicitly. Its general solution is

$$u(\varphi) = \frac{GM}{|\mathbf{L}|^2} + A\cos(\varphi - \varphi_0)$$

where $A, \varphi_0 \in \mathbb{R}$ are constants. After applying a rotation in the $\mathbf{e_1}$-$\mathbf{e_2}$-plane if necessary, we can assume w.l.o.g. that $\varphi_0 = \varphi(t_0) = 0$ and $A \geq 0$. We then have

$$r(t) = \frac{1}{\frac{GM}{|\mathbf{L}|^2} + A\cos(\varphi(t))} = \frac{|\mathbf{L}|^2 / GM}{1 + e \cdot \cos(\varphi(t))},$$

where $e := \frac{A|\mathbf{L}|^2}{GM}$ is called the *eccentricity*. Geometrically, the solution is

(1) an ellipse[1] for $0 \leq e < 1$ (Kepler's first law),
(2) a parabola for $e = 1$,
(3) a hyperbola for $e > 1$.

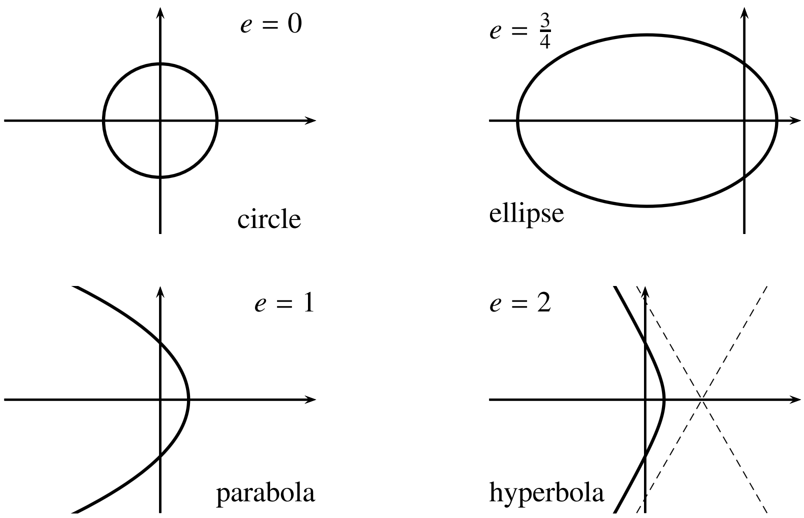


Figure 37.. Kepler orbits

Defining the *gravitational potential* $V : \mathbb{R}^3 \setminus \{\mathbf{0}\} \to \mathbb{R}$ by

$$V(\mathbf{x}) := -\frac{GM}{|\mathbf{x}|}$$

[1] As a special case we get a circle for $e = 0$.

we have

$$-\mathrm{grad}V = -\frac{GM}{r^2}\frac{\partial}{\partial r} = -\frac{GM}{r^3}\mathbf{x} = \frac{1}{m}\mathbf{F}.$$

For the energy we have

$$\begin{aligned} \text{kinetic energy:} \quad & E_{\text{kin}} = \frac{m}{2}|\dot{\mathbf{x}}|^2 \\ \text{potential energy:} \quad & E_{\text{pot}} = m \cdot V(\mathbf{x}) \\ \text{total energy:} \quad & E = E_{\text{kin}} + E_{\text{pot}}. \end{aligned}$$

Lemma 2.7 (Energy Equation). *Let* $\mathbf{x}$ *satisfy* (2.2). *Then*

$$\frac{2}{m}E = \dot{r}^2 + \frac{|\mathbf{L}|^2}{r^2} - \frac{2GM}{r}$$

is constant.

Proof. We have

$$|\dot{\mathbf{x}}|^2 = \left|\dot{r}\frac{\partial}{\partial r} + \dot{\varphi}\frac{\partial}{\partial \varphi}\right|^2 = \dot{r}^2 \cdot 1 + \dot{\varphi}^2 \cdot r^2 = \dot{r}^2 + \frac{|\mathbf{L}|^2}{r^2}$$

because of $\dot{\varphi} = \frac{|\mathbf{L}|}{r^2}$. This implies

$$\frac{2}{m}E = |\dot{\mathbf{x}}|^2 - \frac{2GM}{|\mathbf{x}|} = \dot{r}^2 + \frac{|\mathbf{L}|^2}{r^2} - \frac{2GM}{r}.$$

Therefore

$$\frac{2}{m}\frac{d}{dt}E = 2\dot{r}\ddot{r} - 2\frac{|\mathbf{L}|^2\,\dot{r}}{r^3} + \frac{2GM\dot{r}}{r^2} = 2\dot{r}\Big(\ddot{r} - \frac{|\mathbf{L}|^2}{r^3} + \frac{GM}{r^2}\Big) = 2\dot{r}\Big(\ddot{r} - \dot{\varphi}^2 r + \frac{GM}{r^2}\Big) = 0$$

by the equation of motion (2.4). □

We now define the *effective potential*

$$W(r) := \frac{|\mathbf{L}|^2}{r^2} - \frac{2GM}{r}.$$

By the energy equation $\dot{r}^2 + W(r)$ is constant. From $\dot{r}^2 \geq 0$ we get that $W(r) \leq \text{const}$. The energy diagram is shown in Figure 38..

Remark 2.8. In accordance with (2.1), the graviational force exerted by an extended body with mass distribution ϱ on a point particle with mass m is given by

$$\mathbf{F} = -Gm\int_{\mathbb{R}^3} \frac{\varrho(\mathbf{y})}{|\mathbf{x}-\mathbf{y}|^2}\frac{\mathbf{x}-\mathbf{y}}{|\mathbf{x}-\mathbf{y}|}d\mathbf{y}.$$

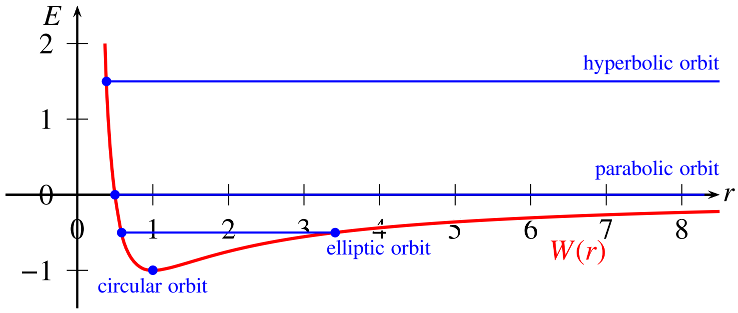


Figure 38.. Energy diagram in Newtonian gravity

2.2. Equivalence principle

Problem. In Newtonian mechanics, moving a heavy mass will *instantaniously* change the gravitational force generated by it everywhere. This can be used to transmit signals with infinite velocity. This contradicts the requirement from special relativity that no information can be tranmitted at a speed higher than that of light. The description of gravity in relativity requires fundamentally new ideas.
Let us go back and investigate what the experimentally confirmed equality of inertial and gravitational mass, $m_{\text{inert}} = m_{\text{grav}}$, tells us. Consider the four situations in depicted in the table below.

A. Our observer is floating in outer space and feels no force. The observer is not accelerated and is at rest in an inertial system.

B. The observer turns on the engine and feels the force generated by the acceleration of the spaceship. The force is proportional to the inertial mass of the observer. Being accelerated, the observer is no longer at rest in an inertial system.

C. The observer is standing still on the surface of the earth. The observer is at rest in an inertial frame but feels the gravitational force of the earth. This force is proportional to the heavy mass of the observer.

D. The spaceship together with the observer is falling freely towards the earth. This is not going to end well but in the meantime the observer feels no force. Being accelerated, the observer is not at rest in an inertial system.
By means of physical experiments, our observer (under complete isolation from the outside world) cannot distinguish situation A from D nor B from C. This known the *equivalence principle* and confirms that $m_{\text{inert}} = m_{\text{grav}}$. On the other hand, A and C are inertial observers, but B and D are not.
Therefore, in a theory incorporating gravity, the observers in A and D should be considered equivalent (similarly for B and C) and we have to give up the concept of inertial frames. Indeed, realistic coordinate systems are usually only local and do not describe the whole universe anyway. They may locally approximate idealized inertial frames. Since we no longer demand existence of global inertial frames we will no longer insist on physical spacetime being modeled by Minkowski space. From now on, spacetime will be modeled by a four-dimensional Lorentz manifold which is not necessarily Minkowski space. The linearization of a Lorentz manifold M

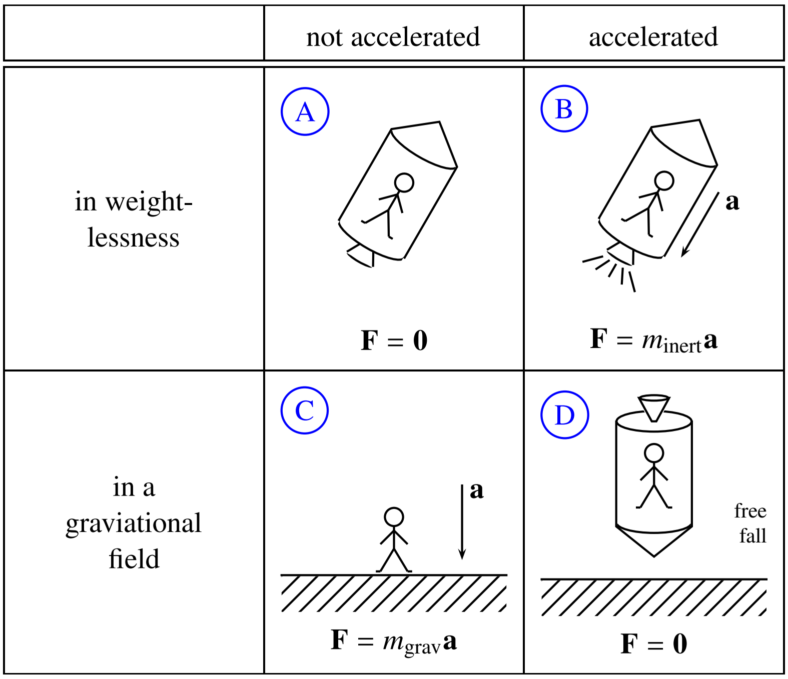


Figure 39.. Equivalence principle

at a point $p \in M$, namely the tangent space T_pM, is isometric to Minkowski space. In this sense special relativity will still be an approximation to general relativity.

2.3. Time orientations

We will now discuss properties of this Lorentz manifold so that it can be considered a reasonable candidate for the mathematical model of spacetime. We are able to distinguish between future and past as we can remember the past but not the future. Mathematically, this is reflected by the concept of time orientation.

Definition 2.9. A *time orientation* on a Lorentz manifold M is a continuous mapping that assigns to each $p \in M$ one of the two connected components of

$$\mathcal{Z}_p := \{v \in T_pM \,|\, g|_p(v, v) < 0\}.$$

Continuity of the time orientation means the follwing: Write $\mathcal{Z}_p^{\uparrow}$ for the connected component selected by the time orientation. We demand that for each $p \in M$ there is an open neighborhood U of p and a continuous vector field v on U such that $v(q) \in \mathcal{Z}_q^{\uparrow}$ for all $q \in U$.

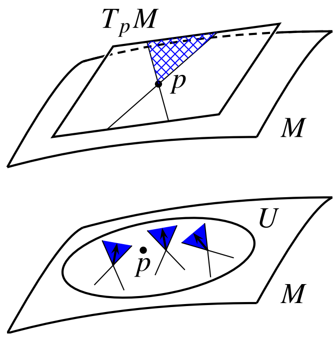
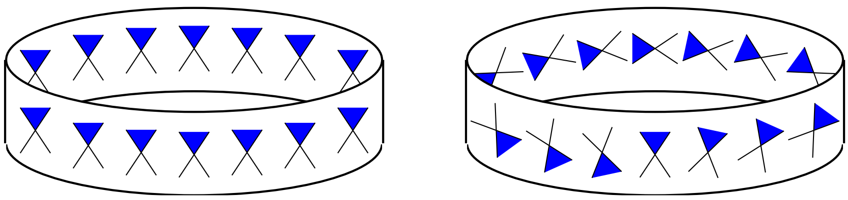


Figure 40.. Time orientation

Example 2.10. In special relativity, we did not have to worry about time orienations because Minkowski space carries a canonical time orienation. Since Minkowski space $\mathbb{R}^4$ is a linear space, each tangent space can be canonically identified with Minkowski space itself. We have implicitly used the constant time orientation given by $p \mapsto \mathcal{Z}^{\uparrow} = \{\mathbf{x} \in \mathbb{R}^4 \mid \langle\langle \mathbf{x}, \mathbf{x} \rangle\rangle < 0 \text{ and } x^0 > 0\}$.

Definition 2.11. A Lorentz manifold is called *time orientable* if has a time orientation. A Lorentz manifold with a selected time orientiation is called *time oriented*.

Example 2.12. Not every Lorentz manifold is time orientable. The following picture shows two different Lorentz metrics on the *same* manifold $\mathbb{R} \times S^1$ so that the first Lorentz manifold is time orientable whereas the second is not.



time orientable not time orientable

Figure 41.. Time orientability

Definition 2.13. Let M be a time oriented Lorentz manifold with time orientation $p \mapsto \mathcal{Z}_p^{\uparrow}$. Timelike tangent vectors $v \in T_pM$ are called *future directed* if $v \in \mathcal{Z}_p^{\uparrow}$ and *past directed* if $-v \in \mathcal{Z}_p^{\uparrow}$. Similarly, lightlike tangent vectors $v \in T_pM$ are called *future directed* if $v \in \partial\mathcal{Z}_p^{\uparrow}$ and *past directed* if $-v \in \partial\mathcal{Z}_p^{\uparrow}$.

Here $\partial\mathcal{Z}_p^{\uparrow}$ denotes the boundary of $\mathcal{Z}_p^{\uparrow}$ in T_pM.

A differentiable curve $c : I \to M$ in a Lorentz manifold will be called timelike if $\dot{c}(t)$ is timelike for all $t \in I$. Lightlike and spacelike curves are defined in the same way. If M is time oriented then we can also talk about future-directed and past-directed timelike or lightlike curves.

2.4. Einstein's field equations

From now on, gravitation will no longer be considered as an external force in contrast to the electromagnetic force, for instance. Instead, it will be modeled by the choice of Lorentz manifold M itself.
In special relativity, the world lines of point particles subject to no force are straight lines. These are the geodesics of Minkowski space. Generalizing this, the Lorentz manifold has to chosen in such a way that the world lines of point particles that move only under the influence of gravitation will be the geodesics of the spacetime.

The big question is now: Which Lorentz manifold should be taken? What is the connection between geometry and physics?

On the physical side we will use the stress-energy tensor T of the matter generating the gravitation. On the geometric side we will use the Einstein tensor G. What is this? Recall the basic curvature tensors of a Lorentz manifold:

1. The *Riemann curvature tensor*
$$R : T_pM \times T_pM \times T_pM \times T_pM \to \mathbb{R}.$$
2. The *Ricci curvature*
$$\mathrm{ric} : T_pM \times T_pM \to \mathbb{R}, \quad \mathrm{ric}(\xi, \eta) := \sum_{i=1}^{n} \varepsilon_i R(\xi, \mathbf{e_i}, \mathbf{e_i}, \eta),$$
where $\mathbf{e_1}, \dots, \mathbf{e_n}$ is a generalized orthonormal basis, that is
$$g(\mathbf{e_i}, \mathbf{e_j}) = \varepsilon_i \delta_{i,j} \text{ with } \varepsilon_i = \pm 1.$$
The map ric is a symmetric bilinear form on T_pM.
3. The *scalar curvature*
$$\mathrm{scal}(p) := \sum_{i=1}^{n} \varepsilon_i \mathrm{ric}(\mathbf{e_i}, \mathbf{e_i}).$$
Then $\mathrm{scal} : M \to \mathbb{R}$ is a function on M.
4. The *Einstein tensor*
$$\mathrm{G} := \mathrm{ric} - \tfrac{1}{2}\mathrm{scal}\cdot g.$$

Why do we use G and not simply ric?

Lemma 2.14. *On any semi-Riemannian manifold we have* $2\mathrm{div}(\mathrm{ric}) = d\mathrm{scal}$ *and hence* $\mathrm{div}(\mathrm{G}) = 0$.

Proof. The divergence of a symmetric $(0,2)$-tensor field like ric is a 1-form defined by

$$\mathrm{div}(\mathrm{ric})(X) = \sum_j \varepsilon_j \nabla_{\mathbf{e_j}} \mathrm{ric}(\mathbf{e_j}, X)$$

where $\mathbf{e_j}$ is a generalized orthonormal tangent frame, $g(\mathbf{e_j}, \mathbf{e_k}) = \varepsilon_j \delta_{jk}$ with $\varepsilon_j = \pm 1$. We now check the formula $2\mathrm{div}(\mathrm{ric}) = d\mathrm{scal}$ at a fixed point p in the manifold and we may assume that X and the tangent frame are synchronous at p, i.e., $\nabla X = \nabla \mathbf{e_j} = 0$ at p. Using the second Bianchi identity we get

$$\begin{aligned}
d\mathrm{scal}(X) &= \partial_X \sum_{jk} \varepsilon_j \varepsilon_k g(R(\mathbf{e_j}, \mathbf{e_k})\mathbf{e_k}, \mathbf{e_j}) \\
&= \sum_{jk} \varepsilon_j \varepsilon_k g(\nabla_X R(\mathbf{e_j}, \mathbf{e_k})\mathbf{e_k}, \mathbf{e_j}) \\
&= -\sum_{jk} \varepsilon_j \varepsilon_k g((\nabla_{\mathbf{e_j}} R(\mathbf{e_k}, X) + \nabla_{\mathbf{e_k}} R(X, \mathbf{e_j}))\mathbf{e_k}, \mathbf{e_j}) \\
&= -\sum_{jk} \varepsilon_j \varepsilon_k (g(\nabla_{\mathbf{e_j}} R(\mathbf{e_k}, X)\mathbf{e_k}, \mathbf{e_j}) + g(\nabla_{\mathbf{e_j}} R(X, \mathbf{e_k})\mathbf{e_j}, \mathbf{e_k})) \\
&= 2\sum_{jk} \varepsilon_j \varepsilon_k g(\nabla_{\mathbf{e_j}} R(\mathbf{e_k}, X)\mathbf{e_j}, \mathbf{e_k}) \\
&= 2\sum_{j} \varepsilon_j \nabla_{\mathbf{e_j}} \mathrm{ric}(X, \mathbf{e_j}) \\
&= 2\mathrm{div}(\mathrm{ric})(X).
\end{aligned}$$

□

The stress-energy tensor T turns out to be divergence free and hence so should be its geometric counterpart. This is the reason for preferring G over ric.
We now postulate the **Einstein field equation.**

$$\boxed{\kappa \cdot T = \mathrm{G}} \tag{EFE}$$

Here κ is a universal constant. The value of κ is determined by transition to the Newtonian limit. If

(1) T is the stress-energy tensor of dust (only mass density),
(2) the Einstein field equation is replaced by its linearization and
(3) c tends to infinity,

then the geodesic equations become Newton's equations of motions with

$$\kappa = \frac{8\pi G}{c^4} \approx 2,07 \cdot 10^{-48} \ \frac{\mathrm{s}^2}{\mathrm{g} \cdot \mathrm{cm}}.$$

It is possible to derive the Einstein field equation from a variational principle and one can give various heuristic arguments for it. Ultimately however, one has to verify it by checking the predicted results experimentally.

Definition 2.15. A Lorentz manifold M is called *vacuum solution* if $T \equiv 0$ and hence (by the Einstein field equation) $\mathrm{G} \equiv 0$.

Example 2.16. Let M be Minkowski space. Here we even have $R \equiv 0$.

Lemma 2.17. *In general, on* 4*-dimensional semi-Riemannian manifolds, we have*

$$\mathrm{ric} = \mathrm{G} - \frac{1}{2}\sum_{i=1}^{4} \varepsilon_i \cdot \mathrm{G}(\mathbf{e_i}, \mathbf{e_i}) \cdot g.$$

Proof. We compute

$$\begin{aligned}\sum_{1=1}^{4} \varepsilon_i \cdot \mathrm{G}(\mathbf{e_i}, \mathbf{e_i}) &= \sum_{i=1}^{4} \varepsilon_i(\mathrm{ric}(\mathbf{e_i}, \mathbf{e_i}) - \tfrac{1}{2}\mathrm{scal} \cdot g(\mathbf{e_i}, \mathbf{e_i}))\\ &= \mathrm{scal} - \tfrac{1}{2}\mathrm{scal} \cdot 4\\ &= -\mathrm{scal}\end{aligned}$$

and therefore

$$\mathrm{G} - \frac{1}{2}\sum_{i=1}^{4} \varepsilon_i \cdot \mathrm{G}(\mathbf{e_i}, \mathbf{e_i}) \cdot g = \mathrm{ric} - \tfrac{1}{2}\mathrm{scal} \cdot g - \tfrac{1}{2}(-\mathrm{scal} \cdot g) = \mathrm{ric}. \qquad \square$$

Corollary 2.18. *We have* $\mathrm{G} = 0$ *if and only if* $\mathrm{ric} = 0$. *Hence by* (EFE) *the vacuum solutions are exactly the Ricci-flat Lorentz manifolds.* $\square$

Since the metric itself is divergence free, i.e., $\mathrm{div}g = 0$, the tensor $G + \Lambda g$ is also divergence free for any constant Λ. Therefore $\mathrm{G} + \Lambda g$ could replace G in the field equation which leads to the **Einstein field equation with cosmological constant**:

$$\mathrm{G} + \Lambda \cdot g = \kappa \cdot T, \qquad (\mathrm{EFE}_\Lambda)$$

where $\Lambda \in \mathbb{R}$ is called the *cosmological constant*. The general opinion whether or not one should allow a nonzero cosmological constant has changed various times. Einstein once considered its introduction as the "greatest stupidity of his life" but changed his mind later. Currently, a nonzero cosmological constant is often considered as necessary to correctly explain the observations.

Example 2.19 (deSitter spacetime). Let $r > 0$. Set

$$\mathrm{S}^4_1(r) := \{\mathbf{x} \in (\mathbb{R}^5, g_{\mathrm{Mink}}) | \langle\langle \mathbf{x}, \mathbf{x} \rangle\rangle = r^2\}.$$

A Lorentz metric is obtained by restricting $\langle\langle \cdot, \cdot \rangle\rangle$ to the tangent spaces of $\mathrm{S}^4_1(r)$. This way, one gets a four-dimensional Lorentz manifold. A time orientation is defined by requiring $x^0 > 0$. A calculation shows

$$\mathrm{G} = -\frac{3}{r^2} g.$$

Hence $\mathrm{S}^4_1(r)$ is a vacuum solution of (EFE_Λ) with $\Lambda = \frac{3}{r^2}$.

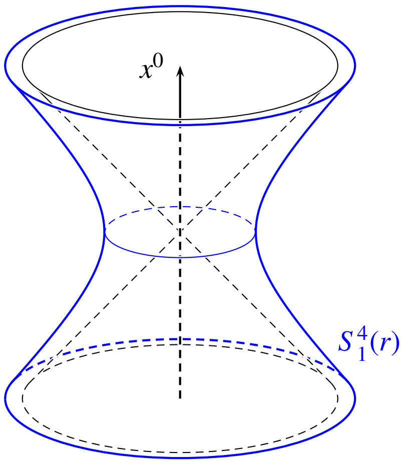


Figure 42.. DeSitter spacetime

Convention. From now on, we will work with physical units chosen in such a way that the speed of light and the gravitational constant are equal to 1. This leads to $\kappa = 8\pi$.

From now on we will accept the **postulate of general relativity**:

Physical spacetime can be identified with a time-oriented 4*-dimensional Lorentz manifold. The world lines of point particles moving only under the influence of gravitation are geodesics, timelike for massive particles and lightlike for particles of rest mass* 0*. The curvature of the manifold and the matter distribution in spacetime are related by the Einstein field equation, possibly with cosmological constant.*

2.5. Exercises

2.1. Explain why Newtonian gravitation theory is incompatible with special relativity.

2.2. Using Newtonian gravitation theory, prove *Kepler's third law* which states that for elliptic orbits,

$$GM \cdot (\text{orbital period})^2 = \frac{4\pi^2}{(1+e)^3} \cdot r_{\text{max}}^3.$$

2.3. Using Newtonian gravitation theory, compute the graviational force of a round homogeneous planet and that of a homogeneous spherical shell.
More precisely, let $\varrho_0 > 0$ and $0 < r < R$. Compute the graviational force vector field for the mass distributions:
(a)

$$\varrho(\mathbf{x}) = \begin{cases} \varrho_0 & \text{if } \|\mathbf{x}\| \leq R, \\ 0 & \text{else.} \end{cases}$$

(b)

$$\varrho(\mathbf{x}) = \begin{cases} \varrho_0 & \text{if } r \leq \|\mathbf{x}\| \leq R, \\ 0 & \text{else.} \end{cases}$$

2.4. On a homogeneous round planet we drill a tunnel from a point on the surface all the way through the center of the planet to the antipodal point on the surface. Compute the trajectory of a point particle falling into the tunnel.
(We assume that the width of the tunnel is so thin that its influence on the graviational field of the planet is negligible. Hence you can use the graviational force you computed in Exercise 2.3.)

2.5. Let $M = \mathbb{R} \times S^1$. Give two explicit examples of Lorentz metrics on M such that the first one is time orientable and second one is not as indicated in Example 2.12. Express them in the coframes dt and $d\theta$ where t is the standard coordinate on $\mathbb{R}$ and θ the angular coordinate on S^1.

2.6. Let $\varrho : \mathbb{R}^4 \to \mathbb{R}$ be the mass density and $\mathbf{u} : \mathbb{R}^4 \to \mathbb{R}^4$ be the unit velocity field of dust on Minkowski space as described in Example 1.30 **(2)**. Let $\mathbf{T}$ be its energy-stress tensor. Show that the following are equivalent:

(i) $\operatorname{div}(\mathbf{T}) = \mathbf{0}$,
(ii) The integral curves of $\mathbf{u}$ (the stream lines of the dust) are straight lines and $\operatorname{div}(\varrho\mathbf{u}) = 0$.

2.7. Let $M > 0$. Show that there is no Riemannian metric on $\mathbb{R}^2 \setminus \{(0,0)^\top\}$ such that its geodesics are precisely Kepler's planetary orbits with central mass M. What if $M = 0$?
Hint: What happens in a Riemannian manifold if two geodesics intersect tangentially?

2.8. Compute the curvature tensor of deSitter space in Example 2.19 using the Gauss equation and verify that deSitter space is a vacuum solution with cosmological constant $\Lambda = \frac{3}{r^2}$.

3. Models for the whole universe

We start by searching Lorentz manifolds which can serve as good models for the whole physical spacetime, i.e. for the whole universe at all times. We will neglect the influence of small-scale structures such as individual stars, galaxies etc. on the geometry of the manifold. In other words, we will pretend that the universe is spatially homogeneous.

3.1. Robertson-Walker spacetimes

Our ansatz is to describe the "spatial part" of the universe by a three-dimensional Riemannian manifold (S, g_S) which is connected and complete, i.e. geodesics are defined for all times. Let us moreover assume that the spatial universe looks the same in each direction, at least when observing objects not too far away. This property is called local isotropy and is formulated mathematically as follows.

Definition 3.1. A Riemannian manifold S is called *locally isotropic*, if for each $p \in S$ and all $X, Y \in T_pS$ with $|X| = |Y|$ there exists an open neighborhood U of p and an isometry $\Phi : U \to U$ with $\Phi(p) = p$ and $d\Phi|_p(X) = Y$.

On a 3-dimensional locally isotropic Riemannian manifold we have for the sectional curvature that $K(E) = K(E')$ whenever $E, E' \subset T_pS$ are 2-dimensional subspaces. The sectional curvature only depends on the base point $p \in M$ but not on the choice of tangential plane over p. This can be seen as follows:
Given planes $E, E' \subset T_pS$ choose $X, Y \in T_pS$ with $|X| = |Y| = 1$ and $X \perp E$ as well as $Y \perp E'$. Then an appropriate local isometry Φ takes X to Y, i.e., $d\Phi|_p(X) = Y$. Hence $d\Phi|_p(E) = E'$. This implies

$$K(E') = K(d\Phi|_p(E)) = K(E)$$

because isometries preserve the curvature.

Theorem 3.2 (Schur). *Let M be a connected Riemannian manifold of dimension $n \geq 3$. If the sectional curvature only depends on the base point then it is constant.*

Proof. Let the sectional curvature depend only on the base point. This means there is a function $\kappa : M \to \mathbb{R}$ such that $K(E) = \kappa(p)$ for all 2-dimensional subspaces $E \subset T_pM$. The Riemann curvature tensor and the sectional curvature contain the same information, they determine each other. In our case this implies

$$R(X, Y, U, V) = \kappa \cdot (g(X, V)g(Y, U) - g(X, U)g(Y, V))$$

where g denotes the metric. Let $X \in T_pM$ be a unit vector. Since the dimension of M is at least 3 we can find $Y, Z \in T_pM$ such that X, Y, Z are orthonormal. The second Bianchi identity implies

$$\begin{aligned}
0 &= \nabla_X R(Y, Z, Z, Y) + \nabla_Y R(Z, X, Z, Y) + \nabla_Z R(X, Y, Z, Y) \\
&= \partial_X \kappa \cdot (g(Y, Y)g(Z, Z) - g(Y, Z)g(Y, Z)) \\
&\quad + \partial_Y \kappa \cdot (g(Z, Y)g(X, Z) - g(Z, Z)g(X, Y)) \\
&\quad + \partial_Z \kappa \cdot (g(X, Y)g(Y, Z) - g(X, Z)g(Y, Y)) \\
&= \partial_X \kappa.
\end{aligned}$$

Since X is arbitrary this implies $d\kappa = 0$ and since the manifold is connected we find that κ is constant. □

Note that Schur's theorem does not hold in dimension 2. The Gauss curvature of a surface is nonconstant in general.

Corollary 3.3. *A connected locally isotropic Riemannian manifold of dimension* 3 *has constant curvature.* □

Remark 3.4. If (N, g) is a Riemannian manifold with sectional curvature K and if $\alpha > 0$, then the Riemannian manifold $(N, c^2 \cdot g)$ has the sectional curvature $\frac{1}{\alpha^2}K$. For this reason it suffices to consider the cases $K \equiv \varepsilon = -1, 0, 1$.

Example 3.5. The prime examples are Euclidean space $(\mathbb{R}^3, g_{\mathrm{eucl}})$ for vanishing curvature ($\varepsilon = 0$), the standard sphere (S^3, g_{std}) for $\varepsilon = 1$ and hyperbolic space (H^3, g_{hyp}) for $\varepsilon = -1$.

Here is a table for the candidates of our Riemannian manifold S for the different choices of ε.

ε	-1	0	1
simply connected	(H^3, g_{hyp})	$(\mathbb{R}^3, g_{\mathrm{eucl}})$	(S^3, g_{std})
quotients	$\vdots$	$T^3 = S^1 \times S^1 \times S^1$ $S^1 \times \mathbb{R}^2$ $T^2 \times \mathbb{R}$ $\vdots$	$\mathbb{R}P^3$ $\vdots$
	infinitely many, not completely understood	essentially finitely many, all known	infinitely many, all known
	some compact, some noncomp. but finite vol., some of infinite volume	some compact, some of infinite vol.	all compact

Conversely, any 3-manifold M of constant curvature ε is locally isometric to (H^3, g_{hyp}), $(\mathbb{R}^3, g_{\mathrm{eucl}})$, or (S^3, g_{std}). Thus every point p has a neighborhood isometric to a ball in one of these three spaces. The group $O(3)$ acts isometrically on the ball letting p fixed. The differentials of the isometries act by the usual action of $O(3)$ on $T_pM \cong \mathbb{R}^3$ and can map any direction to any other direction. Thus all 3-manifolds of constant curvature are locally isotropic.

Now set

$$M := I \times S,$$

for our spacetime, where $I \subset \mathbb{R}$ is an open interval. For the Lorentz metric we make the ansatz

$$\boxed{g = -dt \otimes dt + f(t)^2 \cdot g_S}$$

where $t \in I$ and $f : I \to \mathbb{R}$ is a positive smooth function. Such a metric is called a *warped product*. Put differently: For $\xi = \alpha\frac{\partial}{\partial t} + X, \eta = \beta\frac{\partial}{\partial t} + Y \in T_{(t,p)}M$ with $X, Y \in T_pS$, we have

$$g(\xi, \eta) = -\alpha\beta + f(t)^2 \cdot g_S(X, Y).$$

Definition 3.6. A Lorentz manifold of the form $(M, g) = (I \times S, -dt \otimes dt + f(t)^2 \cdot g_S)$ where (S, g_S) is a 3-dimensional Riemannian manifold of constant sectional curvature will be called a *Robertson-Walker spacetime*.

Example 3.7. For $(S, g_S) = (\mathbb{R}^3, g_{\mathrm{eukl}})$, $I = \mathbb{R}$ and $f = 1$ we obtain the Minkowski space $(M, g) = (\mathbb{R}^4, g_{\mathrm{Mink}})$.

3.1.1. Geodesics of the spacetime

Lemma 3.8. *A curve $s \mapsto c(s) = (t(s), \gamma(s))$ is a geodesic in M if and only if*

(i) $\dfrac{d^2t}{ds^2} + f(t)\dot f(t) g_S(\gamma'(s), \gamma'(s)) = 0$ *and*

(ii) $\dfrac{\nabla^S}{ds}\gamma'(s) + 2\dfrac{\dot f(t)}{f(t)} \cdot \dfrac{dt}{ds} \cdot \gamma'(s) = 0.$

Proof. The condition of being a geodesic is local, so we can check the assertion in local coordinates on S. The metric $-\frac{1}{\varepsilon r^2 - 1} dr \otimes dr + r^2 d\theta \otimes d\theta + r^2 \sin(\theta)^2\, d\varphi \otimes d\varphi$ is readily checked to have constant sectional curvature ε. Since any two Riemannian manifolds of the same constant sectional curvature are locally isometric, we can introduce local coordinates r, θ, φ around any point of S such that $g_S = -\frac{1}{\varepsilon r^2-1} dr \otimes dr + r^2 d\theta \otimes d\theta + r^2 \sin(\theta)^2\, d\varphi \otimes d\varphi$ on that coordinate neighborhood.
Straightforward computation yields for the Christoffel symbols of M with respect to the local coordinates t, r, θ, φ:

$$
\begin{aligned}
\Gamma^t_{rr} &= -\frac{f(t)\dot f(t)}{\varepsilon r^2 - 1} = f(t)\dot f(t)(g_S)_{rr} \\
\Gamma^t_{\theta\theta} &= r^2 f(t)\dot f(t) = f(t)\dot f(t)(g_S)_{\theta\theta} \\
\Gamma^t_{\varphi\varphi} &= r^2 f(t)\sin(\theta)^2\, \dot f(t) = f(t)\dot f(t)(g_S)_{\varphi\varphi}
\end{aligned}
$$

while all other Γ^t_{ij} vanish. This show ((i)).
Comparison of the Christoffel symbols of S and M shows that they coincide for all combinations of spatial coordinates r, θ, φ. Moreover, $\Gamma^r_{tr} = \Gamma^\theta_{t\theta} = \Gamma^\varphi_{t\varphi} = \frac{\dot f(t)}{f(t)}$ and $\Gamma^r_{tt} = \Gamma^\theta_{tt} = \Gamma^\varphi_{tt} = 0$. This yields ((ii)). See the SageMath notebook Robertson-Walker.ipynb for the computations. □

Remark 3.9. Lemma 3.8 holds for all warped products of the form $-dt \otimes dt + f(t)^2 g_S$ whenever (S, g_S) is a Riemannian manifold. The fact that our S is 3-dimensional and has constant sectional curvature is actually irrelevant. See [7, Ch. 7, Prop. 38].

Example 3.10. The curve $c(s) = (s, \gamma_0)$, where $\gamma_0 \in S$ is constant, is a timelike geodesic. We interpret it as the world line of a galaxy.

Let now $s \mapsto c(s) = (t(s), \gamma(s))$ be a null geodesic. Then

$$\begin{aligned}0 &= g\left(\frac{dc}{ds}, \frac{dc}{ds}\right)\\ &= g\left(\frac{dt}{ds}\frac{\partial}{\partial t} + \gamma'(s), \frac{dt}{ds}\frac{\partial}{\partial t} + \gamma'(s)\right)\\ &= -\left(\frac{dt}{ds}\right)^2 + f^2 \cdot g_S(\gamma'(s), \gamma'(s)).\end{aligned}$$

This implies

$$\begin{aligned}\frac{d}{ds}\left(f \cdot \frac{dt}{ds}\right) &= \dot{f} \cdot \left(\frac{dt}{ds}\right)^2 + f \cdot \frac{d^2t}{ds^2}\\ &= \dot{f} \cdot f^2 \cdot g_S(\gamma', \gamma') + f \cdot \frac{d^2t}{ds^2} \overset{\text{(i)}}{=} 0.\end{aligned}$$

Hence $f \cdot \frac{dt}{ds}$ is constant.

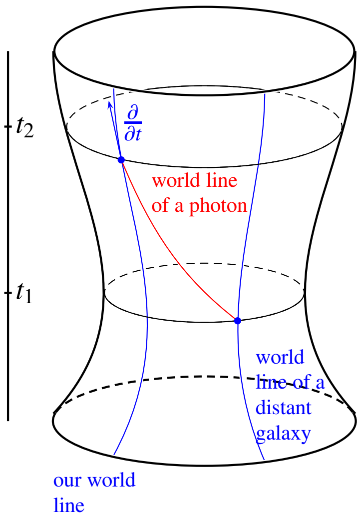


Figure 43.. Redshift in Robertson-Walker spacetime

From the point of view of an observer with world line $s \mapsto (s, \gamma_0)$, we get for the energy of a photon

$$E = g\left(\frac{\partial}{\partial t}, \frac{dc}{ds}\right) = g\left(\frac{\partial}{\partial t}, \frac{dt}{ds}\frac{\partial}{\partial t} + \gamma'(s)\right) = -\frac{dt}{ds} = -\frac{\text{const}}{f(t)}$$

and therefore

$$\frac{E(t_1)}{E(t_2)} = \frac{f(t_2)}{f(t_1)}.$$

Definition 3.11. The quantity

$$z := \frac{f(t_2) - f(t_1)}{f(t_1)} = \frac{f(t_2)}{f(t_1)} - 1$$

is called *redshift* (of the null geodesic).

Redshift can be observed and measured very precisely because chemical elements emit light at specific energy levels (frequencies, colors). This light, emitted by other galaxies arrives with a color shift which can be measured to high precision.
Taylor expansion of f in the variable t_2 yields

$$\begin{aligned}f(t_1) &= f(t_2) + \dot{f}(t_2)(t_1 - t_2) + O(|t_1 - t_2|^2)\\ &= f(t_2)(1 + H(t_2)(t_1 - t_2) + O(|t_1 - t_2|^2)),\end{aligned}$$

where $H(t) = \frac{\dot{f}(t)}{f(t)}$. This implies

$$
\begin{aligned}
z &= \frac{1}{1 + H(t_2)(t_1 - t_2) + O(|t_1 - t_2|^2)} - 1 \\
&= 1 - H(t_2)(t_1 - t_2) + O(|t_1 - t_2|^2) + O(|t_1 - t_2|^2) - 1 \\
&= H(t_2)(t_2 - t_1) + O(|t_1 - t_2|^2).
\end{aligned}
$$

Hence if we observe light from galaxies not too far away, such that the term $O(|t_1 - t_2|^2)$ is negligible compared to the term $H(t_2)(t_2 - t_1)$, then the redshift is essentially proportional to the time difference $|t_1 - t_2|$, hence to the distance to the other galaxy. The constant $H_0 = H(t_0)$ where t_0 stands for "now" is called the *Hubble constant*. In fact, one observes $z > 0$, so the Hubble constant is positive. Therefore $\dot{f}$(now) is positive, i.e., the universe is currently expanding.

Lemma 3.12. *The Ricci curvature of a Robertson-Walker spacetimes satisfies*

$$
\begin{aligned}
\mathrm{ric}\left(\frac{\partial}{\partial t}, \frac{\partial}{\partial t}\right) &= -3\frac{\ddot{f}}{f}, \\
\mathrm{ric}\left(\frac{\partial}{\partial t}, X\right) &= \mathrm{ric}\left(X, \frac{\partial}{\partial t}\right) = 0, \\
\mathrm{ric}(X, Y) &= \left\{2\left(\frac{\dot{f}}{f}\right)^2 + 2\frac{\varepsilon}{f^2} + \frac{\ddot{f}}{f}\right\} g(X, Y),
\end{aligned}
$$

where X and Y are tangent to S.

Proof. Use either explicit computation (see SageMath notebook Robertson-Walker.ipynb) or the general formulas for warped products (see [7, Ch. 7, Cor. 43]). ◻

This implies

$$
\mathrm{scal} = 3\frac{\ddot{f}}{f} + 3\left\{2\left(\frac{\dot{f}}{f}\right)^2 + 2\frac{\varepsilon}{f^2} + \frac{\ddot{f}}{f}\right\} = 6\left\{\left(\frac{\dot{f}}{f}\right)^2 + \frac{\varepsilon}{f^2} + \frac{\ddot{f}}{f}\right\}.
$$

For the Einstein tensor we find

$$
\begin{aligned}
G\left(\frac{\partial}{\partial t}, \frac{\partial}{\partial t}\right) &= 3\left\{\left(\frac{\dot{f}}{f}\right)^2 + \frac{\varepsilon}{f^2}\right\}, \\
G\left(\frac{\partial}{\partial t}, X\right) &= G\left(X, \frac{\partial}{\partial t}\right) = 0, \\
G(X, Y) &= -\left\{\left(\frac{\dot{f}}{f}\right)^2 + \frac{\varepsilon}{f^2} + 2\frac{\ddot{f}}{f}\right\} g(X, Y).
\end{aligned}
\tag{1}
$$

If the Einstein field equations (with $\Lambda = 0$) are satisfied then the momentum density in the energy-stress tensor must vanish and the stress density must be isotropic. We get

$$\frac{8\pi}{3}\varrho = \left(\frac{\dot f}{f}\right)^2 + \frac{\varepsilon}{f^2}, \tag{2}$$

$$-8\pi p = \left(\frac{\dot f}{f}\right)^2 + \frac{\varepsilon}{f^2} + 2\frac{\ddot f}{f}, \tag{3}$$

where ϱ is the energy density and p the pressure.
Subtracting (2) from (3) gives

$$3\frac{\ddot f}{f} = -4\pi(\varrho + 3p). \tag{4}$$

Differentiation of (2) and insertion of (4) and (2) yields

$$\begin{aligned}\frac{8\pi}{3}\dot\varrho &= 2 \cdot \frac{\dot f}{f} \cdot \frac{\ddot f f - \dot f^2}{f^2} - 2 \cdot \frac{\varepsilon \dot f}{f^3} \\ &= \left(2\frac{\ddot f}{f} - 2\left(\left(\frac{\dot f}{f}\right)^2 + \frac{\varepsilon}{f^2}\right)\right) \cdot \frac{\dot f}{f} \\ &= \left(-\frac{8\pi}{3}(\varrho + 3p) - \frac{16\pi}{3}\varrho\right) \cdot \frac{\dot f}{f} \\ &= (-8\pi\varrho - 8\pi p) \cdot \frac{\dot f}{f},\end{aligned}$$

hence

$$\dot\varrho = -3(\varrho + p) \cdot \frac{\dot f}{f}. \tag{5}$$

3.1.2. Singularities

Let the domain $I = (t_*, t^*)$ of f be maximal in the sense that f cannot be extended beyond I as a positive smooth function. Here $-\infty \leq t_* < t^* \leq \infty$.

Definition 3.13. (1) t_* or t^* is called a *physical singularity*, if $\varrho \to \infty$ for $t \searrow t_*$ or $t \nearrow t^*$, respectively.
(2) t_* is called a *big bang*, if $f(t) \to 0$ and $\dot f(t) \to \infty$ for $t \searrow t_*$.
(3) t^* is called a *big crunch* or *collapse*, if $f(t) \to 0$ and $\dot f(t) \to -\infty$ for $t \nearrow t^*$.

Remark 3.14. If $\varrho + 3p \geq 0$ and $H_0 = H(t_0) > 0$, then M has an initial singularity, i.e.,

$t_* > -\infty$. To see this, notice that f is concave because $\ddot{f} = -\frac{4\pi}{3}(\varrho + 3p)f \leq 0$. From $H(t_0) > 0$ we see $\dot{f}(t_0) > 0$ and, by concavity, $\dot{f} \geq \dot{f}(t_0)$ on $(t_*, t_0]$. This implies that for any $t_1 \in (t_*, t_0]$,

$$f(t_0) > f(t_0) - f(t_1) = \int_{t_1}^{t_0} \dot{f}(t)\, dt \geq (t_0 - t_1) \cdot \dot{f}(t_0)$$

and thus

$$t_0 - t_1 \leq \frac{f(t_0)}{\dot{f}(t_0)} = \frac{1}{H_0}.$$

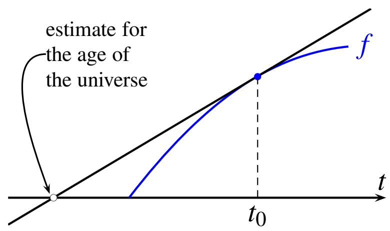


Figure 44.. Initial singularity

If we let t_1 tend to t_* we obtain

$$t_0 - t_* \leq \frac{1}{H_0}.$$

This way, we did not only show $t_* > -\infty$, but also derived an upper bound for the age of the universe in terms of the Hubble constant. Remember that the Hubble constant can be quite well determined experimentally via the observation of redshift. Current estimates for the age of the universe give it a value of about 13.8 billion years.

Proposition 3.15. *Suppose t_* and t^* are physical singularities if they are finite. Let $H_0 > 0$, $\varrho > 0$ and suppose there are constants $-\frac{1}{3} < a < A$, such that $a \leq \frac{p}{\varrho} \leq A$. Then*

(1) The initial singularity t_* is a big bang.
(2) If $\varepsilon = 0$ or $\varepsilon = -1$, then $I = (t_*, \infty)$ and $f \to \infty$, $\varrho \to 0$ for $t \to \infty$.
(3) If $\varepsilon = 1$, then $I = (t_*, t^*)$ and $t^* < \infty$ is a big crunch.

Proof. Set $\delta := 3a + 1$. Because of $a > -\frac{1}{3}$ we have $\delta > 0$ and

$$\begin{aligned} -\frac{1}{3} + \frac{\delta}{3} = a \leq \frac{p}{\varrho} \quad &\Longrightarrow \quad 3p + \varrho \geq \delta\varrho > 0 \\ &\overset{(4)}{\Longrightarrow} \quad \ddot{f} < 0 \\ \big(\text{with } \dot{f}(t_0) > 0\big) \quad &\Longrightarrow \quad \dot{f} > 0 \text{ on } (t_*, t_0]. \end{aligned}$$

Furthermore, we have on $(t_*, t_0]$

$$\dot{\varrho} \overset{(5)}{=} -3(\varrho + p)\frac{\dot{f}}{f} \geq -3(\varrho + A\varrho)\frac{\dot{f}}{f} = -C \cdot \varrho \cdot \frac{\dot{f}}{f}$$

with $C := 3(1 + A) > 0$. This implies

$$\begin{aligned}(\ln\varrho)^{\cdot} \geq -\,C(\ln f)^{\cdot} = (\ln f^{-C})^{\cdot} \\ \implies \quad & (\ln(\varrho f^{C}))^{\cdot} \geq 0 \\ \implies \quad & (\varrho f^{C})^{\cdot} \geq 0 \\ \implies \quad & \varrho \cdot f^{C} \leq \varrho(t_0) f(t_0)^{C} \quad \text{on} \quad (t_*, t_0] \\ \implies \quad & f \to 0 \quad \text{for} \quad t \to t_*\end{aligned}$$

because $\varrho \nearrow \infty$ as $t \searrow t_*$. Moreover, we have

$$\dot{\varrho} \overset{(5)}{=} -3(\varrho + p)\frac{\dot{f}}{f} \leq -(\delta\varrho + 2\varrho)\frac{\dot{f}}{f} = -(2+\delta)\varrho \cdot \frac{\dot{f}}{f}$$

and in a similar fashion to before we obtain

$$(\varrho f^{2+\delta})^{\cdot} \leq 0 \quad \implies \quad \underbrace{\varrho \cdot f^{2+\delta}}_{=\varrho f^{2}\cdot f^{\delta}} \geq \varrho(t_0) f(t_0)^{2+\delta} \text{ on } (t_*, t_0].$$

With $f \to 0$, we get $f^{\delta} \to 0$ and thus $\varrho f^{2} \to \infty$ for $t \to t_*$. With (2) we then get

$$\infty \leftarrow \frac{8\pi}{3}\varrho f^{2} = \dot{f}^{2} + \varepsilon \quad \implies \quad \dot{f} \to \infty.$$

This shows (1).
Case 1: The function f has a maximum at $t_m \in I$.

$$0 < \varrho(t_m) \overset{(2)}{=} \frac{3}{8\pi}\Big\{\overbrace{\frac{\dot{f}(t_m)^2}{f(t_m)^2}}^{=0} + \frac{\varepsilon}{f(t_m)^2}\Big\} \quad \implies \quad \varepsilon > 0, \quad \text{which means} \quad \varepsilon = 1.$$

With $\ddot{f} < 0$ this implies $\dot{f} < 0$ on (t_m, t^*). A discussion similar to the one before shows that t^* is a big crunch. This is the situation in (3).
Case 2: The function f does not have a maximum on I.
This implies $\dot{f} > 0$ on I because $\ddot{f} < 0$. From $\varrho > 0$ and $3p + \varrho > 0$ it follows that $3(p + \varrho) > 0$. Hence

$$\dot{\varrho} = -3(\varrho + p)\frac{\dot{f}}{f} < 0.$$

Thus I does not have an ending physical singularity, $t^* = \infty$.
Subcase A: $f \to \infty$ for $t \to \infty$.
We have $(\varrho f^{2+\delta})^{\cdot} \leq 0$ so that $\varrho f^{2+\delta}$ is bounded on (t_0, ∞). With $f \to \infty$, this implies $\varrho f^{2} \to 0$ for $t \to \infty$. Now (2) implies

$$0 \leftarrow \frac{8\pi}{3}\varrho f^{2} \overset{(2)}{=} \dot{f}^{2} + \varepsilon \quad \implies \quad \varepsilon \leq 0$$

i.e., $\varepsilon = -1$ or $\varepsilon = 0$. This is the situation in (2).

Subcase B: $f \to b < \infty$ for $t \to \infty$.
This implies $\dot{f} \to 0$ for $t \to \infty$.

$$\implies \quad \frac{8\pi}{3}\varrho f^2 = \dot{f}^2 + \varepsilon \xrightarrow{t\to\infty} \varepsilon \quad \implies \quad \varepsilon \geq 0$$

which means $\varepsilon = 0$ or $\varepsilon = 1$.

Since ϱf^C is positive and increasing, we have $\varrho f^2 \nrightarrow 0$. This implies $\varepsilon \neq 0$, i.e. $\varepsilon = 1$. This shows $\varrho f^2 \to \frac{3}{8\pi}$ for $t \to \infty$.
On the other hand, by the mean value theorem, there is a sequence $t_i \in (i, i-1)$ with $\ddot{f}(t_i) = \dot{f}(i+1) - \dot{f}(i)$. Because of $\dot{f} \to 0$ we then get $\ddot{f}(t_i) \to 0$.

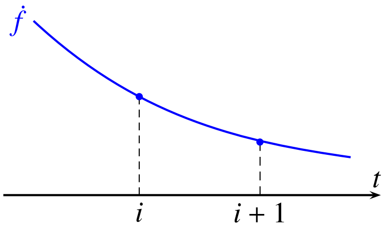


Figure 45.. Decay of $\ddot{f}$

This implies

$$3\frac{\ddot{f}}{f} \overset{(4)}{=} -4\pi(\varrho + 3p) \quad \implies \quad \varrho(t_i) + 3p(t_i) \to 0 \text{ for } i \to \infty.$$

Now,

$$0 \leftarrow \varrho(t_i) + 3p(t_i) \geq \delta\varrho(t_i) \quad \implies \quad \varrho(t_i) \to 0 \quad \implies \quad \varrho(t_i)f(t_i)^2 \to 0.$$

This is a contradiction, so Subcase B does not occur. □

Remark 3.16. In the literature, the case $\varepsilon \leq 0$ is often called *open* case as opposed to the *closed* case $\varepsilon > 0$, because if S is simply connected and complete then S is noncompact for $\varepsilon \leq 0$ ($S = \mathbb{R}^3$ or $S = H^3$) and S is compact for $\varepsilon > 0$ ($S = S^3$). Indeed, S is always compact when $\varepsilon = 1$. However, when $\varepsilon = 0$ or $\varepsilon = -1$, the manifold S can be compact as well, for example for $\varepsilon = 0$ we could have $S = T^3$. Therefore this terminology is somewhat misleading.

Definition 3.17. The constant $\varrho_c := \dfrac{3{H_0}^2}{8\pi}$ is called the *critical energy density*.

The reason for this terminology is given by

Proposition 3.18. *We have*

$$\begin{aligned} \varrho(t_0) < \varrho_c \quad &\iff \quad \varepsilon = -1 \\ \varrho(t_0) = \varrho_c \quad &\iff \quad \varepsilon = 0 \\ \varrho(t_0) > \varrho_c \quad &\iff \quad \varepsilon = 1 \end{aligned}$$

Proof. From the equation (2) we find

$$\varrho(t_0) - \varrho_c = \frac{3}{8\pi}\left\{H_0{}^2 + \frac{\varepsilon}{f(t_0)^2} - H_0{}^2\right\} = \frac{3}{8\pi}\cdot\frac{\varepsilon}{f(t_0)^2}.$$

Thus the sign of ε is the same as that of $\varrho(t_0) - \varrho_c$. $\square$

Definition 3.19. A *dust cosmos* is a Robertson-Walker spacetime with $p = 0$. A dust cosmos with $H_0 > 0$ is called *Friedmann cosmos*.

Proposition 3.20. *For a Robertson-Walker spacetime M with nonconstant f, the following statements are equivalent.*

(i) M is a dust cosmos.
(ii) $\varrho \cdot f^3 =: m$ is constant.
(iii) The Friedmann equation

$$\dot{f}^2 + \varepsilon = \frac{A}{f}$$

holds with the constant $A = \frac{8\pi}{3}m > 0$.

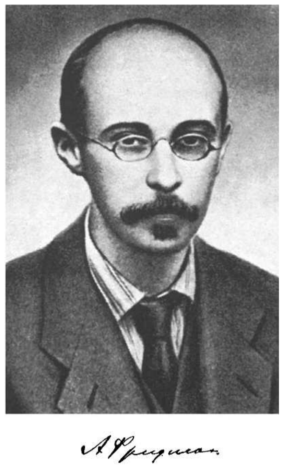


Figure 46.. Alexander A. Friedmann (1888–1925)

Proof. “(ii) ⇔ (iii)” is clear because by (2) we have

$$\frac{\frac{8\pi}{3}\varrho f^3}{f} = \dot{f}^2 + \varepsilon.$$

“(i)⇒(ii)”: If $p = 0$ equation (5) yields $\dot{\varrho} = -3\varrho\frac{\dot{f}}{f}$ which implies $(\ln\varrho)^\cdot + 3(\ln f)^\cdot = 0$. Hence $(\ln\varrho f^3)^\cdot = 0$ and thus $(\varrho f^3)^\cdot = 0$.
“(ii)⇒(i)” On the one hand, $(\varrho f^3)^\cdot = 0$ implies

$$\dot{\varrho} = -3\varrho\frac{\dot{f}}{f}.$$

On the other hand, we have by (5)

$$\dot{\varrho} = -(3\varrho + p)\frac{\dot{f}}{f}.$$

Therefore

$$p \cdot \dot{f} = 0.$$

Now set $J := \{t \in I \mid p(t) \neq 0\}$. We have to show $J = \emptyset$. Suppose $J \neq \emptyset$ and let J_0 be a nonempty connected component of J. Then J_0 is an open interval. It follows that $\dot{f} \equiv 0$ on J_0, hence $f \equiv a > 0$ on J_0. By (3) this means

$$-8\pi p \equiv \frac{\varepsilon}{a^2} \neq 0 \quad \text{on} \quad J_0.$$

Since p is continuous,

$$p \equiv -\frac{\varepsilon}{8\pi a^2} \neq 0 \quad \text{on} \quad \overline{J_0}$$

where $\overline{J_0}$ is the closure of J_0 in I. This implies $\overline{J_0} \subset J$, i.e., $J_0 = \overline{J_0}$. Thus $J_0 = I$. Hence f is constant, in contradiction to the assumption. □

We now determine the solutions of the Friedmann equation. Without loss of generality, let $t_* = 0$.

(1) $\varepsilon = 0$: $f(t) = (\frac{3}{2})^{\frac{2}{3}} A^{\frac{1}{3}} t^{\frac{2}{3}}$ (semicubical parabola).
(2) $\varepsilon = -1$: The function $[0,\infty) \to [0,\infty)$, $T \mapsto \frac{A}{2}(\sinh(T) - T)$ is strictly increasing. Hence for any $t \geq 0$ we can uniquely solve $t = \frac{A}{2}(\sinh(T) - T)$ for T. Putting $f = \frac{A}{2}(\cosh(T) - 1)$ we check

$$\dot{f}^2 + \varepsilon = \left(\frac{\sinh(T)}{\cosh(T) - 1}\right)^2 - 1 = \frac{2}{\cosh(T) - 1} = \frac{A}{f}.$$

(3) $\varepsilon = 1$: Similar reasoning as above yields

$$t = \tfrac{A}{2}(T - \sin(T)) \text{ and } f = \tfrac{A}{2}(1 - \cos(T)).$$

This describes a cycloid.

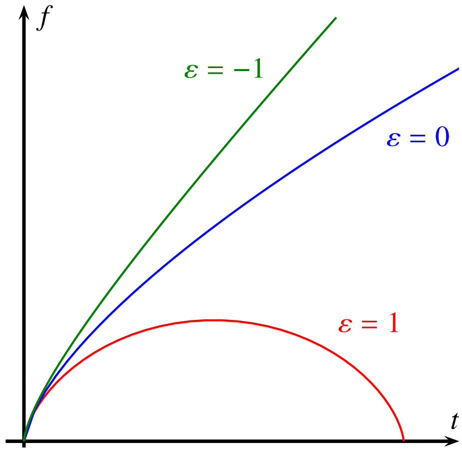


Figure 47.. Solutions to the Friedmann equation

Definition 3.21. A Robertson-Walker spacetime is called a *radiation cosmos*, if

$$p = \frac{\varrho}{3}.$$

In Exercise 3.5 it will be shown that

(1) $\varrho \cdot f^4 =: A$ is constant.

(2) We have $f(t)^2 = -\varepsilon(t-t_*)^2 + 4\sqrt{\frac{2\pi}{3}A} \cdot (t-t_*)$.

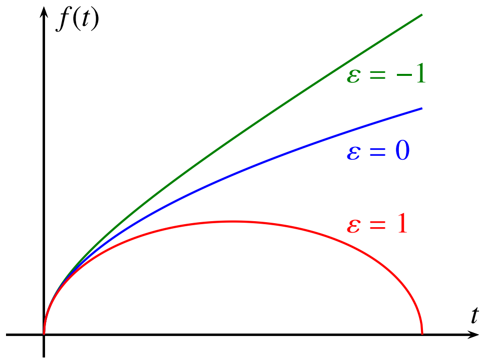


Figure 48. Radiation cosmos

Remark 3.22. The existence of big bang singularities is not that much dependent on the particular ansatz used here but can be derived in great generality. This is the content of Hawking's *singularity theorem*, see e.g. [7, Ch. 14].

3.1.3. Horizons

Let $M = I \times S$ be a Robertson-Walker spacetime with distortion function f. Let $\gamma : [s_0, \infty) \to M$,

$$\gamma(s) = (\underbrace{\gamma^0(s)}_{\in I}, \underbrace{\hat{\gamma}(s)}_{\in S})$$

a future-directed null curve such as the world line of a photon. Then

$$0 = -((\gamma^0)')^2 + f(\gamma^0)^2 \cdot \left\|\hat{\gamma}'\right\|_S^2 \quad \Longrightarrow \quad \left\|\hat{\gamma}'\right\|_S = \frac{(\gamma^0)'}{f(\gamma^0)}.$$

For the length of $\hat{\gamma}$ we have

$$\int_{s_0}^{\infty} \left\|\hat{\gamma}'(s)\right\|_S \, ds = \int_{s_0}^{\infty} \frac{(\gamma^0)'(s)}{f(\gamma^0(s))} \, ds = \int_{\gamma^0(s_0)}^{\infty} \frac{d\gamma^0}{f(\gamma^0)}.$$

If f growth fast enough, for example $f(t) = t^2$ or $f(t) = e^t$ we find

$$R := \int_{s_0}^{\infty} \left\|\hat{\gamma}'(s)\right\|_S \, ds < \infty. \tag{6}$$

This shows that photons starting at a point p cannot leave the ball around p of radius R in S. This means that parts of the universe cannot be observed. This is known as the *horizon problem*.

3.2. Cosmological inflation

We consider a Robertson Walker spacetime $M = I \times S$, where $I \subset \mathbb{R}$ is an interval, (S, g_S) is a Riemannian manifold of dimension 3 and M is equipped with the metric

$$g = -dt^2 + f(t)^2 g_S, \quad \text{where } f : I \to \mathbb{R}$$

is positive and smooth. Our aim is to find spacetimes of this form which provide a better explanation of some astronomical observations than the models considered in the previous section.

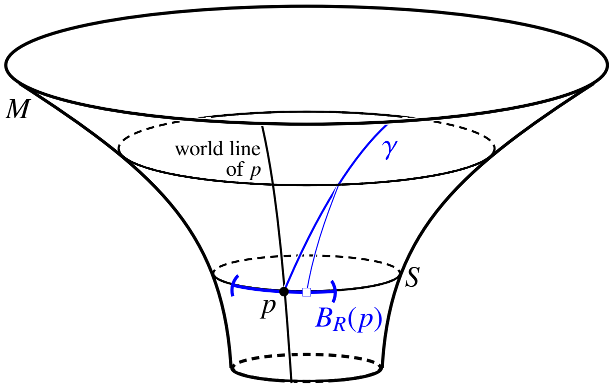


Figure 49.. Horizon problem

3.2.1. Two problems with Friedmann spacetimes

We describe two observational facts which can be better explained if the cosmic expansion function f grows exponentially during a period shortly after the big bang. As a first example we mention that the energy density ρ and the critical energy density ρ_c of our universe can be estimated using observations by astronomers. They find that $|\frac{\rho}{\rho_c} - 1| \leq 0.04$ at the present time. In order to estimate this quantity at times shortly after the big bang we compute, using (2),

$$\frac{\rho}{\rho_c} - 1 = \frac{H^2 + \varepsilon f^{-2}}{H^2} - 1 = \frac{\varepsilon}{f^2 H^2} = \frac{\varepsilon}{\dot{f}^2} \quad \text{where } H = \frac{\dot{f}}{f}.$$

Since $t = 0$ is a big bang singularity we have $\dot{f}(t) \nearrow \infty$ as $t \searrow 0$. Thus

$$\frac{\rho}{\rho_c} - 1 = \frac{\varepsilon}{\dot{f}^2} \to 0, \quad \text{as } t \searrow 0.$$

This means that shortly after the big bang ρ was even closer to ρ_c than it is today. In other words: By this model, our universe can have evolved into its present state only if a very special initial condition was satisfied. This is not very likely to happen in nature unless there is a reason for it. Therefore we ask: Why should ρ and ρ_c be so close to each other? We note that this can be explained by an exponential expansion of the universe shortly after the big bang. Indeed if there is an interval $I' \subset I$ such that for all $t \in I'$ we have $f(t) = a \exp(\lambda t)$ for some $a, \lambda > 0$ then we get on I':

$$\frac{\rho}{\rho_c} - 1 = \frac{\varepsilon}{a^2 \lambda^2 \exp(2\lambda t)}.$$

This function decays exponentially and thus at the end of I' the quantity $\frac{\rho}{\rho_c} - 1$ is very small even if it was large at the beginning.

As a second problem, we mention that the theory of inflation can also explain better why the cosmic microwave background radiation is almost the same in all directions. We noted in Section 3.1.3 that a photon which is emitted at $t = t_0$ and received at $t = t_1$ can cover distance in S at most

$$r = \int_{t_0}^{t_1} \frac{dt}{f(t)}.$$

In the following we take t_1 to be our time, i.e. 13.7×10^9 years after the big bang and we let t_0 be the time when the cosmic microwave background radiation which we observe was emitted, i.e. 3.8×10^5 years after the big bang. Earth is a point $E\in S$ and we define the 2-sphere of radius r around E in S by

$$S^2(E,r) := \{x\in S \mid \mathrm{dist}_S(x,E) < r\}.$$

It is a remarkable fact that the cosmic microwave background radiation which we observe is distributed very uniformly in all directions. How can we explain this fact?

First we note that under the assumption that f grows polynomially it is very hard to give a physical reason. In order to see this, recall that the solution of the Friedmann equation for $\varepsilon = 0$ is given by $f(t) = a\cdot t^{2/3}$ where a is positive constant. The solutions for $\varepsilon = \pm 1$ still satisfy $f(t)\sim a\cdot t^{2/3}$ as $t\searrow 0$ because $t\sim \frac{A}{12}T^3$ and $f\sim \frac{A}{4}T^2$. So consider $f(t) = a\cdot t^{2/3}$. Then the distance in S traced out by the photon between time t_0 and t_1 is bounded by

$$\int_{t_0}^{t_1}\frac{dt}{at^{2/3}} = \tfrac{3}{a}(t_1^{1/3}-t_0^{1/3}) =: r(t_0,t_1).$$

The photons of the cosmic microwave background radiation which we observe were emitted from $S^2(E,r(t_0,t_1))$. Now note that

$$\frac{r(t_0,t_1)}{r(0,t_0)} = \frac{t_1^{1/3}-t_0^{1/3}}{t_0^{1/3}} = \Big(\frac{t_1}{t_0}\Big)^{1/3} - 1 \approx 16.$$

Thus there are configurations consisting of many points in $S^2(E,r(t_0,t_1))$ with mutual distance larger than $r(0,t_0)$, see also Exercises 3.7 and 3.8. Two points in $B(E,r(t_0,t_1))$ with distance larger than $r(0,t_0)$ cannot have influenced each other during the time interval $(0,t_0)$. So why is the cosmic microwave background radiation so uniform on $S^2(E,r(t_0,t_1))$?

On the other hand, if f grows exponentially, $f(t) = a\exp(\lambda t)$ for $t\geq t'$, then

$$\int_{t'}^{t}\frac{d\tau}{a\exp(\lambda\tau)} = \tfrac{1}{a\lambda}(\exp(-\lambda t')-\exp(-\lambda t)) =: \widetilde{r}(t',t)$$

is bounded as $t\to\infty$. Therefore in this case one can expect a more uniform distribution of radiation.

3.2.2. A simple inflationary model

We now describe a simple inflationary model which leads to a period of exponential expansion of the universe. We consider an unknown scalar field $\varphi\colon M\to\mathbb{R}$ on M, the so called *inflaton field* and we use the Lagrangian density

$$\mathcal{L}_{\mathrm{infl}}(\varphi,g) := -(\langle d\varphi, d\varphi\rangle_g + 2V(\varphi))d\mathrm{vol}_g$$

where $V\colon\mathbb{R}\to\mathbb{R}$ is a given potential. In order to obtain an equation for φ we use the Lagrangian principle, i.e. we demand that φ is critical for $\mathcal{L}_{\mathrm{infl}}(\varphi,g)$. More precisely, for all compact subsets

$K \subset M$ and for all smooth 1-parameter families $\varphi_s = \varphi + sv$, $v \in C_c^\infty(M)$, $\mathrm{supp}(v) \subset K$, we require

$$\frac{d}{ds}\Big|_{s=0} \int_K \mathcal{L}_{\mathrm{infl}}(\varphi_s, g) = 0.$$

We get

$$\begin{aligned} 0 &= -\frac{d}{ds}\Big|_{s=0} \int_K (\langle d\varphi + sdv, d\varphi + sdv\rangle_g + 2V(\varphi + sv))d\mathrm{vol}_g \\ &= -2\int_K (\langle d\varphi, dv\rangle_g + V'(\varphi)v)d\mathrm{vol}_g \\ &= -2\int_K (\Box\varphi + V'(\varphi))v\, d\mathrm{vol}_g, \end{aligned}$$

where $\Box := -\mathrm{div} \circ \mathrm{grad}$ is the d'Alembert operator. Since this needs to hold for all v the inflaton field must solve a wave equation:

$$\Box\varphi + V'(\varphi) = 0. \tag{7}$$

In order to determine the energy-momentum tensor T_{infl} of the inflaton we use again the Lagrangian principle: For all compact subsets $K \subset M$ and for all smooth 1-parameter families $g_s = g + sh$, $\mathrm{supp}(h) \subset K$, we require

$$\frac{d}{ds}\Big|_{s=0} \int_K \mathcal{L}_{\mathrm{infl}}(\varphi, g_s) = \int_K \langle 8\pi T_{\mathrm{infl}}, h\rangle\, d\mathrm{vol}_g.$$

Using that

$$\frac{d}{ds}\Big|_{s=0} d\mathrm{vol}_{g_s} = \tfrac{1}{2}\langle g, h\rangle\, d\mathrm{vol}_g$$

and that

$$\frac{d}{ds}\Big|_{s=0} \langle d\varphi, d\varphi\rangle = \frac{d}{ds}\Big|_{s=0} g_s^{ij}\partial_i\varphi\partial_j\varphi = -g^{ik}g^{j\ell}h_{k\ell}\partial_i\varphi\partial_j\varphi = -\langle d\varphi \otimes d\varphi, h\rangle_g$$

we obtain

$$\begin{aligned} \int_K \langle 8\pi T_{\mathrm{infl}}, h\rangle_g\, d\mathrm{vol}_g &= -\int_K (\langle d\varphi, d\varphi\rangle_g + 2V(\varphi))\tfrac{1}{2}\langle g, h\rangle\, d\mathrm{vol}_g + \int_K \langle d\varphi \otimes d\varphi, h\rangle_g\, d\mathrm{vol}_g \\ &= \int_K \Big\langle d\varphi \otimes d\varphi - \Big(\tfrac{1}{2}\langle d\varphi, d\varphi\rangle_g + V(\varphi)\Big)g, h\Big\rangle\, d\mathrm{vol}_g \end{aligned}$$

and thus

$$8\pi T_{\mathrm{infl}} = d\varphi \otimes d\varphi - (\tfrac{1}{2}\langle d\varphi, d\varphi\rangle + V(\varphi))g.$$

We write down the Einstein field equations $G = 8\pi T_{\mathrm{infl}}$ using the results (1) for the Einstein tensor of g:

$$\begin{aligned} 3\Big(H^2 + \frac{\varepsilon}{f^2}\Big) = G(\partial_t, \partial_t) &= 8\pi T_{\mathrm{infl}}(\partial_t, \partial_t) = \dot\varphi^2 + \tfrac{1}{2}\langle d\varphi, d\varphi\rangle + V(\varphi), \\ 0 = G(\partial_t, X) &= 8\pi T_{\mathrm{infl}}(\partial_t, X) = \dot\varphi\partial_X\varphi \end{aligned}$$

and with $X \in TS$, $g(X,X) = 1$, we get

$$-\Big(H^2 + \frac{\varepsilon}{f^2} + 2\frac{\ddot{f}}{f}\Big) = G(X,X) = 8\pi T_{\text{infl}}(X,X) = (\partial_X\varphi)^2 - \Big(\tfrac{1}{2}\langle d\varphi, d\varphi\rangle + V(\varphi)\Big).$$

By spatial homogeneity we assume that $\partial_X\varphi = 0$ for all $X \in TS$. Then we have $d\varphi = \dot{\varphi}dt$ and thus the Einstein field equations read

$$\begin{aligned} 3\Big(H^2 + \frac{\varepsilon}{f^2}\Big) &= \dot{\varphi}^2 + \tfrac{1}{2}(-\dot{\varphi}^2) + V(\varphi) = \tfrac{1}{2}\dot{\varphi}^2 + V(\varphi),\\ -\Big(H^2 + \frac{\varepsilon}{f^2} + 2\frac{\ddot{f}}{f}\Big) &= \tfrac{1}{2}\dot{\varphi}^2 - V(\varphi) \end{aligned}$$

We calculate

$$\dot{H} = \frac{\ddot{f}f - \dot{f}^2}{f^2} = \frac{\ddot{f}}{f} - H^2$$

and thus the Einstein field equations read

$$\begin{aligned} &3H^2 + 3\frac{\varepsilon}{f^2} = \tfrac{1}{2}\dot{\varphi}^2 + V(\varphi),\\ &\tfrac{1}{2}\dot{\varphi}^2 - V(\varphi) = -\Big(H^2 + \frac{\varepsilon}{f^2} + 2(\dot{H} + H^2)\Big) = -\Big(2\dot{H} + \frac{\varepsilon}{f^2} + 3H^2\Big). \end{aligned}$$

By adding and subtracting these two equations we get

$$\dot{H} + 2\frac{\varepsilon}{f^2} + 3H^2 = V(\varphi) \tag{8}$$

$$\dot{H} - \frac{\varepsilon}{f^2} = -\tfrac{1}{2}\dot{\varphi}^2. \tag{9}$$

Since we have $\partial_X\varphi = 0$ for all $X \in TS$ the wave equation (7) for φ reads

$$\ddot{\varphi} + 3H\dot{\varphi} + V'(\varphi) = 0. \tag{10}$$

From now we assume that $\varepsilon = 0$ for simplicity. Then the Einstein field equations read

$$\dot{H} + 3H^2 = V(\varphi) \tag{11}$$

$$\dot{H} = -\tfrac{1}{2}\dot{\varphi}^2. \tag{12}$$

In particular we have $\dot{H} \leq 0$, i.e. H is monotonically decreasing.

Example 3.23. The simplest example for the inflaton potential is clearly $V \equiv 0$. The general solution to the equation $\dot{H} + 3H^2 = 0$ is given by

$$H(t) = \begin{cases} \frac{1}{3t+a}, & \text{if } H(0) = \frac{1}{a},\\ 0, & \text{if } H(0) = 0, \end{cases}$$

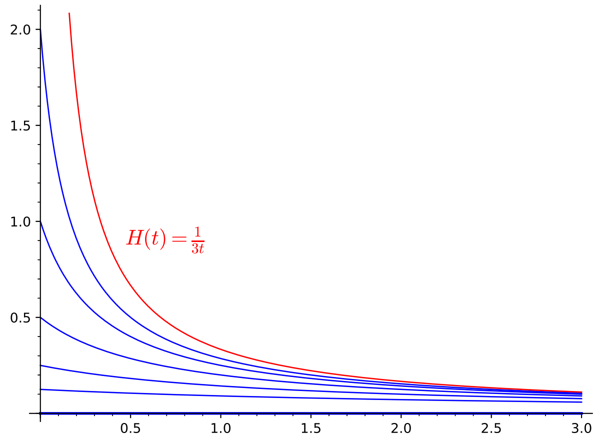


Figure 50.. Hubble constant for $V = 0$

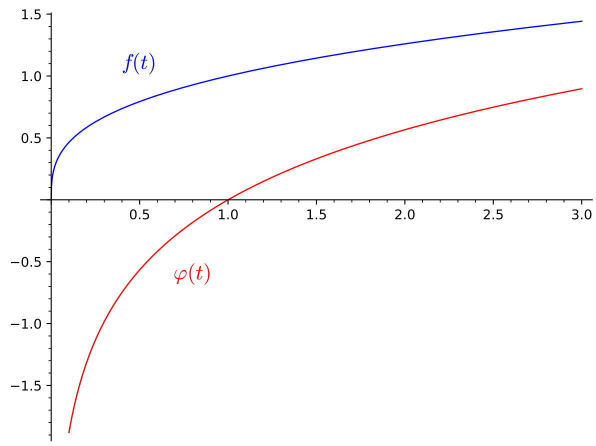


Figure 51.. Expansion factor and inflaton for $V = 0$

where $a \neq 0$.
Since we are looking for a solution with a big bang, we require $H(t) \to \infty$ as $t \searrow 0$ and thus we take $H(t) = \frac{1}{3t}$. Since we have $H(t) = \frac{d}{dt}\ln(f(t))$ we get $f(t) = c_1 t^{1/3}$ with $c_1 \in \mathbb{R}$. In particular we have $\dot{f}(t) = \frac{c_1}{3}t^{-2/3} \to \infty$ as $t \searrow 0$ and thus we have found a solution with a big bang singularity. We also have $\dot{\varphi}^2 = -2\dot{H} = \frac{2}{3t^2}$ and thus $\varphi(t) = \pm\sqrt{\frac{2}{3}}\ln(t) + c_2$ with $c_2 \in \mathbb{R}$.

Example 3.24. Now consider $V \equiv v_0 > 0$. The general solution to the equation $\dot{H} + 3H^2 = v_0$

for $H(0) > 0$ is given by

$$H(t) = \begin{cases} \sqrt{\frac{v_0}{3}} \coth(\sqrt{3v_0}(t+a)), & \text{if } H(0) = \sqrt{\frac{v_0}{3}} \coth(\sqrt{3v_0}a) > \sqrt{\frac{v_0}{3}}, \\ \sqrt{\frac{v_0}{3}}, & \text{if } H(0) = \sqrt{\frac{v_0}{3}}, \\ \sqrt{\frac{v_0}{3}} \tanh(\sqrt{3v_0}(t+a)), & \text{if } H(0) = \sqrt{\frac{v_0}{3}} \tanh(\sqrt{3v_0}a) < \sqrt{\frac{v_0}{3}}. \end{cases}$$

where $a \in \mathbb{R}$. Since we are looking for a solution with a big bang, we require $H(t) \to \infty$ as $t \searrow 0$, and thus we take $H(t) = \sqrt{\frac{v_0}{3}} \coth(\sqrt{3v_0}t)$.

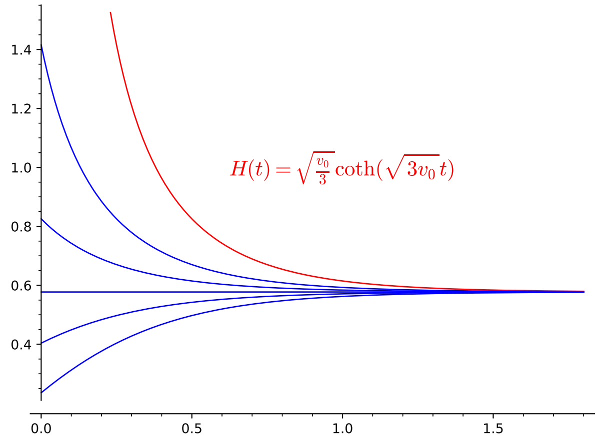


Figure 52.. Hubble constant for $V = v_0$

In order to find f we compute

$$-\frac{d}{dt}\frac{1}{6}\ln\Big(\frac{3}{v_0}H^2 - 1\Big) = -\frac{1}{6}\frac{3v_0^{-1}\cdot 2H\dot{H}}{3v_0^{-1}H^2 - 1} = -\frac{1}{6}\frac{6H(v_0 - 3H^2)}{3H^2 - v_0} = H = \frac{d}{dt}\ln f$$

and thus

$$f(t) = c_1\Big(\frac{3}{v_0}H^2 - 1\Big)^{-1/6} = c_1\Big(\coth^2(\sqrt{3v_0}t) - 1\Big)^{-1/6} = c_1\Big(\sinh(\sqrt{3v_0}t)\Big)^{1/3},$$

where $c_1 > 0$. In particular as $t \searrow 0$ we get $f(t) \searrow 0$ and

$$\begin{aligned} \dot{f}(t) &= \frac{c_1}{3}\Big(\sinh(\sqrt{3v_0}t)\Big)^{-2/3}\sqrt{3v_0}\cosh(\sqrt{3v_0}t) \\ &= c_1\sqrt{\frac{v_0}{3}}\Big(\sinh(\sqrt{3v_0}t)\Big)^{-2/3}\cosh(\sqrt{3v_0}t) \to \infty. \end{aligned}$$

Hence we again found a solution with a big bang singularity. Note that in this example the function f is exponentially growing as $t \to \infty$. We also have

$$\dot{\varphi}^2 = -2\dot{H} = -2(v_0 - 3H^2) = -2\Big(v_0 - v_0\coth^2(\sqrt{3v_0}t)\Big) = \frac{2v_0}{\sinh^2(\sqrt{3v_0}t)}.$$

Thus we get

$$\varphi(t) = \pm\sqrt{\tfrac{2}{3}}\ln\Big(\tanh\Big(\frac{\sqrt{3v_0}}{2}t\Big)\Big) + c_2$$

with $c_2 \in \mathbb{R}$.

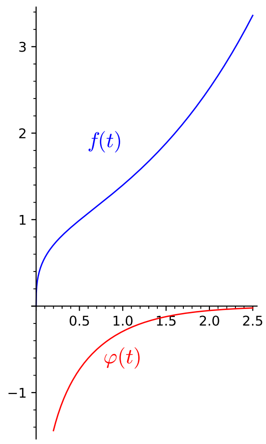


Figure 53.. Expansion factor and inflaton for $V = v_0$

In order to estimate solutions to the field equations (11), (12) we will use the following comparison lemma.

Lemma 3.25. *Let $I \subset \mathbb{R}$ be an open interval, let $h_0, h_1\colon I \to \mathbb{R}$ be C^1 functions such that $\dot{h}_0 + 3h_0^2 \leq \dot{h}_1 + 3h_1^2$ and let $t_0 \in I$.*

(1) If $h_0(t_0) \leq h_1(t_0)$ then we have $h_0(t) \leq h_1(t)$ for all $t \in I$, $t \geq t_0$.
(2) If $h_0(t_0) \geq h_1(t_0)$ then we have $h_0(t) \geq h_1(t)$ for all $t \in I$, $t \leq t_0$.

Proof. We define $F\colon I \to \mathbb{R}$ by

$$F(t) := (h_1(t) - h_0(t))\exp\Big(3\int_{t_0}^{t}(h_1(\tau) + h_0(\tau))d\tau\Big).$$

We compute

$$\begin{aligned}\dot{F}(t) &= (\dot{h}_1(t) - \dot{h}_0(t))\exp\Big(3\int_{t_0}^{t}(h_1(\tau) + h_0(\tau))d\tau\Big)\\ &\quad + (h_1(t) - h_0(t))3(h_1(t) + h_0(t))\exp\Big(3\int_{t_0}^{t}(h_1(\tau) + h_0(\tau))d\tau\Big)\\ &= (\dot{h}_1(t) - \dot{h}_0(t) + 3h_1(t)^2 - 3h_0(t)^2)\exp\Big(3\int_{t_0}^{t}(h_1(\tau) + h_0(\tau))d\tau\Big)\end{aligned}$$

$$\geq 0.$$

Hence the function F is monotonically increasing. If $h_0(t_0) \leq h_1(t_0)$ then $F(t_0) \geq 0$ and thus for all $t \geq t_0$ we have $F(t) \geq 0$ i.e. $h_1(t) \geq h_0(t)$. This shows the first assertion and the second one follows in the same way. □

Next we construct a potential V and functions H, φ satisfying the Einstein field equations (11), (12) such that there exist t_0, t_1 with $0 < t_0 < t_1$ such that f grows exponentially in $(0, t_0)$ and f grows polynomially in (t_1, ∞).

Example 3.26. Let $v_0, \varphi_0 > 0$. We define the following constants

$$\begin{aligned}
t_0 &:= \tfrac{2}{\sqrt{3v_0}}\operatorname{artanh}\Big(\exp\Big(-\sqrt{\tfrac{3}{2}}\varphi_0\Big)\Big), \\
c_1 &:= \sqrt{\tfrac{3}{v_0}}\tanh\big(\sqrt{3v_0}t_0\big) - 3t_0, \\
c_2 &:= \frac{\sqrt{v_0}}{\sinh(\sqrt{3v_0}t_0)}, \\
t_1 &:= \tfrac{1}{3}\Big(\sqrt{\tfrac{6}{c_2^2+2v_0}} - c_1\Big), \\
\varphi_1 &:= \varphi_0 - c_2(t_1 - t_0) < \varphi_0.
\end{aligned}$$

Now let $V\colon \mathbb{R} \to \mathbb{R}$ be such that $0 \leq V \leq v_0$ everywhere, $V(\varphi) = v_0$ for all $\varphi \geq \varphi_0$ and $V(\varphi) = 0$ for all $\varphi \leq \varphi_1$.

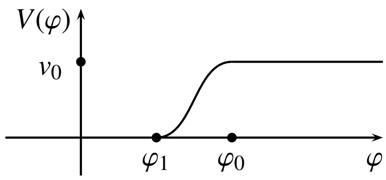


Figure 54.. Inflationary potential

From Example 3.24 we know that for t close to 0 the functions

$$H(t) = \sqrt{\tfrac{v_0}{3}}\coth\big(\sqrt{3v_0}t\big), \quad \varphi(t) = -\sqrt{\tfrac{2}{3}}\ln\Big(\tanh\Big(\tfrac{\sqrt{3v_0}}{2}t\Big)\Big)$$

solve the equations (11), (12) provided that

$$\varphi(t) \geq \varphi_0, \quad \text{i.e.} \quad t \leq \tfrac{2}{\sqrt{3v_0}}\operatorname{artanh}\Big(\exp\Big(-\sqrt{\tfrac{3}{2}}\varphi_0\Big)\Big) = t_0.$$

Thus on $[0, t_0]$ the function f coincides with the one in Example 3.24 and grows exponentially.

In particular, we have $\varphi(t_0) = \varphi_0$. Using Lemma 3.25 with $h_0 = H$, $h_1 = \sqrt{\frac{v_0}{3}}\coth(\sqrt{3v_0}t)$, $V \leq v_0$, we conclude that for all t we have $H(t) \leq \sqrt{\frac{v_0}{3}}\coth(\sqrt{3v_0}t)$. Using Lemma 3.25 with $h_0 = \frac{1}{3t+c_1}$, $h_1 = H$, $V \geq 0$, we conclude that for $t \geq t_0$ we have

$$H(t) \geq \frac{1}{3t+c_1}. \tag{13}$$

Here we have chosen c_1 such that

$$H(t_0) = \frac{1}{3t+c_1}, \quad \text{i.e.} \quad c_1 = \sqrt{\frac{3}{v_0}}\tanh\left(\sqrt{3v_0}t_0\right) - 3t_0.$$

By (13) and (11) we have for all $t \geq t_0$

$$\frac{3}{(3t+c_1)^2} \leq 3H(t)^2 = \tfrac{1}{2}\dot{\varphi}(t)^2 + V(\varphi(t)) \leq \tfrac{1}{2}\dot{\varphi}(t)^2 + v_0$$

and thus

$$\dot{\varphi}(t)^2 \geq \frac{6}{(3t+c_1)^2} - 2v_0. \tag{14}$$

In particular, we have

$$\dot{\varphi}(t_0)^2 \geq \frac{6}{(3t_0+c_1)^2} - 2v_0 = 6\cdot\frac{v_0}{3}\coth^2\left(\sqrt{3v_0}t_0\right) - 2v_0 = \frac{2v_0}{\sinh^2(\sqrt{3v_0}t_0)} = 2c_2^2.$$

By continuity of $\dot{\varphi}$, there exists $t_1 > t_0$ such that for all $t \in [t_0, t_1]$ we have

$$\dot{\varphi}(t)^2 \geq c_2^2.$$

By (14) and since the map $t \mapsto \frac{6}{(3t+c_1)^2} - 2v_0$ is monotonically decreasing, we can find a possible t_1 with this property by requiring

$$\frac{6}{(3t_1+c_1)^2} - 2v_0 = c_2^2, \quad \text{i.e.} \quad t_1 = \tfrac{1}{3}\Big(\sqrt{\tfrac{6}{c_2^2+2v_0}} - c_1\Big).$$

Therefore we have $|\dot{\varphi}| \geq c_2$ on $[t_0, t_1]$ and since $\dot{\varphi}(t_0) < 0$ we have $\dot{\varphi} < -c_2$ on $[t_0, t_1]$. It follows that

$$\varphi(t_1) = \varphi(t_0) + \int_{t_0}^{t_1}\dot{\varphi}(\tau)d\tau \leq \varphi_0 - c_2(t_1 - t_0) = \varphi_1$$

and thus $V(\varphi(t_1)) = 0$. For all $t \geq t_1$ the functions

$$H(t) = \frac{1}{3t+c_3}, \quad \varphi(t) = -\sqrt{\tfrac{2}{3}}\ln\left(t + \tfrac{c_3}{3}\right) + c_4$$

solve the equations (11), (12) by Example 3.23 where c_3, c_4 are determined by

$$H(t_1) = \frac{1}{3t_1+c_3}, \quad \varphi(t_1) = -\sqrt{\tfrac{2}{3}}\ln\left(t_1 + \tfrac{c_3}{3}\right) + c_4.$$

Namely since $V(\varphi(t_1)) = 0$ and φ is monotonically decreasing we have $V(\varphi(t)) = 0$ for all $t \geq t_1$. The function f coincides with the one in Example 3.23 up to a time shift.

3.3. Exercises

3.1. Show that deSitter spacetime $S^4_1(1)$ is a Robertson-Walker spacetime. What are S, g_S, f and ε in this case?

3.2. Let (M, g) be a Robertson-Walker spacetime which is a vacuum solution with trivial cosmological constant $\Lambda = 0$.
(a) Show that (M, g) is flat, i.e., the whole curvature tensor vanishes.
(b) Give an example with $\varepsilon = -1$.

3.3. Determine the solutions of the Friedmann equation for general ε, i.e. without normalizing to $\varepsilon \in \{-1, 0, 1\}$.

3.4. We scale the Friedmann cosmos such that the coordinate t gives the age of the universe in years. For this we have to give up the normalization $\varepsilon \in \{-1, 0, 1\}$.
(a) Today ($t = 13.700.000.000$) the energy density ρ coincides with the critical energy density ρ_c up to an error of 4 percent. Deduce an upper bound on $|\varepsilon|$ from this.
(b) What was the error 380.000 years after big bang?

3.5. Show that for any radiation cosmos the following holds:
(a) $\varrho \cdot f^4 =: A$ is constant.
(b) We have $f(t)^2 = -\varepsilon(t - t_*)^2 + 4\sqrt{\frac{2\pi}{3}A} \cdot (t - t_*)$.

3.6. Let $\gamma_1(s) = (s, p_1)$ and $\gamma_2(s) = (s, p_2)$ be the world lines of two galaxies in a Robertson-Walker spacetime with warping function $f(t) = e^t$. Show that the galaxies move away from one another with more than the speed of light. Why does this not lead to causality problems as tachyons do in special relativity?

3.7. Let $0 < r < R$. We put $S^2(R) := \{x \in \mathbb{R}^3 \mid \|x\| = R\}$. Let $\alpha(r, R)$ be the maximal number of points on $S^2(R)$ which have pairwise Euclidean distance r at least. The exercise aims at a proof of the estimate

$$\frac{3}{4}\frac{R^2}{r^2} + \frac{1}{4} \leq \alpha(r, R) \leq 24\frac{R^2}{r^2} + 2.$$

(a) Prove the upper bound.
Hint: Check that the Euclidean balls of radius $\frac{r}{2}$ about these points are pairwise disjoint. Therefore the sum of their volumes is not larger than the volume of the $\frac{r}{2}$-neighborhood of $S^2(R)$.
(b) Show that for a maximal system of points as above, the balls radius $2r$ cover the r-neighborhood of $S^2(R)$.
(c) Use this to derive the lower bound.

3.8. We consider the spacetime $M = (0, \infty) \times \mathbb{R}^3$ with $g = -dt^2 + f(t)^2 g_{\mathrm{eukl}}$. Use the results of Exercise 3.7 to derive an estimate on the number of causally independent directions (at time

$t = 380.000$), from which we can observe the cosmic background microwave radiation today ($t = 13.700.000.000$) where
(a) $f(t) = a \cdot t^{2/3}$,
(b) $f(t) = \exp(t)$.

3.9. (a) Show that in the inflation model the wave equation (10) for the inflaton field φ follows from the Einstein field equations (8) and (9) provided $\dot{\varphi} \neq 0$.
(b) Show that if $\varepsilon \geq 0$ and $V < 0$ in the inflation model, then each solution satisfies $\dot{\varphi}(t) \neq 0$ for all t.

4. Black holes

In order to find relativistic planetary orbits analogous to Kepler's orbits we need to find Lorentzian manifolds modeling the spacetime outside a gravitating object. The simplest one is the Schwarzschild spacetime which turns out to be a good model for a static, non-rotating radially symmetric astronomical object. It comes with one parameter which corresponds to the mass of the gravitating object. The model predicts interesting phenomena not observed in Newtonian gravitation, namely the existence of a horizon. Once something crosses the horizon towards the central mass, it can never return, not even light. This leads to the terminology of a black hole.
If one wants to allow rotating central masses one has to generalize the Schwarzschild model and allow for a second parameter encoding the rotation speed. This is then known as the Kerr solution. All these solutions model spacetime away from the central mass and are therefore vacuum solutions, i.e. they are Ricci-flat.

4.1. The Schwarzschild solution

In order to find a model for a vacuum spacetime outside a static, radially symmetric astronomical object, we make the following ansatz:
Set $M := \mathbb{R} \times J \times S^2$, where $J \subset \mathbb{R}$, and for $t \in \mathbb{R}$, $\tilde{r} \in J$ set

$$g := -F(\tilde{r})^2 dt \otimes dt + H(\tilde{r})^2 d\tilde{r} \otimes d\tilde{r} + G(\tilde{r})^2 g_{S^2}$$

with positive smooth functions $F, G, H : J \to \mathbb{R}$. Here t is the parameter in the first $\mathbb{R}$-factor and $\tilde{r} \in J$. The $\mathbb{R}$-factor is to be thought of as the time direction and $J \times S^2$ generalizes 3-dimensional Euclidean space in polar coordinates. Since the metric is supposed to be radially symmetric, the functions F, G, and H are independent of the variables parametrizing S^2 and since it is static they are also independent of t.

Figure 55.. Karl Schwarzschild (1873–1916)

Without loss of generality, we assume $H \equiv 1$ because otherwise we may substitute $\tilde{\tilde{r}} = h(\tilde{r})$ with $h' = H$.

After introducing polar coordinates φ, ϑ on S^2 its metric takes the form

$$g_{S^2} = \sin^2\vartheta d\varphi \otimes d\varphi + d\vartheta \otimes d\vartheta.$$

Hence

$$\begin{aligned} g &= -F(\tilde r)^2 dt \otimes dt + d\tilde r \otimes d\tilde r \\ &\quad + G(\tilde r)^2(\sin^2\vartheta d\varphi \otimes d\varphi + d\vartheta \otimes d\vartheta). \end{aligned}$$

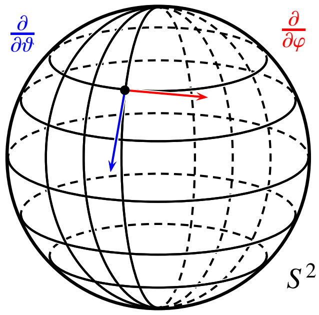


Figure 56.. Polar coordinates on S^2

The mapping $(t, \tilde r, \varphi, \vartheta) \mapsto (2t_0 - t, \tilde r, \varphi, \vartheta)$ is an isometry. This implies that the fixed point set $\{t_0\} \times J \times S^2 =: N_1(t_0)$ is a totally geodesic hypersurface. Its unit normal field is given by

$$\nu_1 = \frac{1}{F(\tilde r)}\frac{\partial}{\partial t}.$$

Since $N_1(t_0)$ is totally geodesic, we have $\nabla_\xi \nu_1 = 0$ for all ξ tangent to $N_1(t_0)$ and thus

$$\begin{aligned} \nabla_{\frac{\partial}{\partial\tilde r}}\frac{\partial}{\partial t} &= \nabla_{\frac{\partial}{\partial\tilde r}}((F(\tilde r)\nu_1) = F'(\tilde r)\frac{1}{F(\tilde r)}\frac{\partial}{\partial t} + 0 = \frac{F'(\tilde r)}{F(\tilde r)}\frac{\partial}{\partial t}, \\ \nabla_{\frac{\partial}{\partial\varphi}}\frac{\partial}{\partial t} &= \nabla_{\frac{\partial}{\partial\varphi}}\frac{\partial}{\partial t} = 0. \end{aligned}$$

The mapping $(t, \tilde r, \varphi, \vartheta) \mapsto (t, \tilde r, 2\varphi_0 - \varphi, \vartheta)$ is an isometry as well, so once again, its fixed point set $\mathbb{R} \times J \times \{\sigma \in S^2 | \varphi(\sigma) = \varphi_0\} =: N_2(\varphi_0)$ is a totally geodesic hypersurface. In this case, its unit normal field is given by

$$\nu_2 = \frac{1}{G(\tilde r)\sin(\vartheta)}\frac{\partial}{\partial\varphi}.$$

Once again, for all ξ tangent to $N_2(\varphi_0)$, we have $\nabla_\xi \nu_2 = 0$ and

$$\begin{aligned} \nabla_{\frac{\partial}{\partial\tilde r}}\frac{\partial}{\partial\varphi} &= \frac{G'(\tilde r)}{G(\tilde r)}\frac{\partial}{\partial\varphi}, \\ \nabla_{\frac{\partial}{\partial\vartheta}}\frac{\partial}{\partial\varphi} &= \cot(\vartheta)\frac{\partial}{\partial\varphi}. \end{aligned}$$

For the covariant derivative of $\frac{\partial}{\partial\varphi}$ in direction $\frac{\partial}{\partial\varphi}$, therefore we get

$$\begin{aligned} \left\langle \nabla_{\frac{\partial}{\partial\varphi}}\frac{\partial}{\partial\varphi}, \frac{\partial}{\partial\varphi}\right\rangle &= \tfrac{1}{2}\frac{\partial}{\partial\varphi}\underbrace{\left\langle\frac{\partial}{\partial\varphi}, \frac{\partial}{\partial\varphi}\right\rangle}_{=G(\tilde r)^2\sin(\vartheta)^2} = 0, \\ \left\langle \nabla_{\frac{\partial}{\partial\varphi}}\frac{\partial}{\partial\varphi}, \frac{\partial}{\partial t}\right\rangle &= \frac{\partial}{\partial\varphi}\underbrace{\left\langle\frac{\partial}{\partial\varphi}, \frac{\partial}{\partial t}\right\rangle}_{=0} - \left\langle\frac{\partial}{\partial\varphi}, \underbrace{\nabla_{\frac{\partial}{\partial\varphi}}\frac{\partial}{\partial t}}_{=0}\right\rangle = 0, \\ \left\langle \nabla_{\frac{\partial}{\partial\varphi}}\frac{\partial}{\partial\varphi}, \frac{\partial}{\partial\tilde r}\right\rangle &= -\left\langle\frac{\partial}{\partial\varphi}, \nabla_{\frac{\partial}{\partial\varphi}}\frac{\partial}{\partial\tilde r}\right\rangle = -\left\langle\frac{\partial}{\partial\varphi}, \frac{G'}{G}\frac{\partial}{\partial\varphi}\right\rangle = -G'G\sin(\vartheta)^2, \\ \left\langle \nabla_{\frac{\partial}{\partial\varphi}}\frac{\partial}{\partial\varphi}, \frac{\partial}{\partial\vartheta}\right\rangle &= -\left\langle\frac{\partial}{\partial\varphi}, \nabla_{\frac{\partial}{\partial\varphi}}\frac{\partial}{\partial\vartheta}\right\rangle = -\cot(\vartheta)\left\langle\frac{\partial}{\partial\varphi}, \frac{\partial}{\partial\varphi}\right\rangle := -\sin(\vartheta)\cos(\vartheta)G^2. \end{aligned}$$

This shows

$$\nabla_{\frac{\partial}{\partial\varphi}}\frac{\partial}{\partial\varphi} = -G'G\sin(\vartheta)^2\frac{\partial}{\partial\tilde{r}} - \sin(\vartheta)\cos(\vartheta)\frac{\partial}{\partial\vartheta}.$$

The other covariant derivatives of the coordinate fields can be derived in a similar fashion. Collecting all derivatives we have

$$\begin{aligned}
\nabla_{\frac{\partial}{\partial\tilde{r}}}\frac{\partial}{\partial t} = \nabla_{\frac{\partial}{\partial t}}\frac{\partial}{\partial\tilde{r}} &= \frac{F'}{F}\frac{\partial}{\partial t},\\
\nabla_{\frac{\partial}{\partial\varphi}}\frac{\partial}{\partial t} = \nabla_{\frac{\partial}{\partial t}}\frac{\partial}{\partial\varphi} &= 0,\\
\nabla_{\frac{\partial}{\partial\vartheta}}\frac{\partial}{\partial t} = \nabla_{\frac{\partial}{\partial t}}\frac{\partial}{\partial\vartheta} &= 0,\\
\nabla_{\frac{\partial}{\partial\tilde{r}}}\frac{\partial}{\partial\varphi} = \nabla_{\frac{\partial}{\partial\varphi}}\frac{\partial}{\partial\tilde{r}} &= \frac{G'}{G}\frac{\partial}{\partial\varphi},\\
\nabla_{\frac{\partial}{\partial\vartheta}}\frac{\partial}{\partial\varphi} = \nabla_{\frac{\partial}{\partial\varphi}}\frac{\partial}{\partial\vartheta} &= \cot(\vartheta)\frac{\partial}{\partial\varphi},\\
\nabla_{\frac{\partial}{\partial\vartheta}}\frac{\partial}{\partial\vartheta} &= -G'G\frac{\partial}{\partial\tilde{r}},\\
\nabla_{\frac{\partial}{\partial\varphi}}\frac{\partial}{\partial\varphi} &= -\sin(\vartheta)^2G'G\frac{\partial}{\partial\tilde{r}} - \sin(\vartheta)\cos(\vartheta)\frac{\partial}{\partial\vartheta},\\
\nabla_{\frac{\partial}{\partial\tilde{r}}}\frac{\partial}{\partial\tilde{r}} &= 0,\\
\nabla_{\frac{\partial}{\partial t}}\frac{\partial}{\partial t} &= F'F\frac{\partial}{\partial\tilde{r}}.
\end{aligned}$$

Since we are looking for a vacuum solution we equate the Ricci curvature to 0 and get

$$0 \overset{!}{=} \operatorname{ric}\left(\frac{\partial}{\partial t},\frac{\partial}{\partial t}\right) = F\left(F'' + 2F'\frac{G'}{G}\right), \tag{1}$$

$$0 \overset{!}{=} \operatorname{ric}\left(\frac{\partial}{\partial\tilde{r}},\frac{\partial}{\partial\tilde{r}}\right) = -\left(\frac{F''}{F} + 2\frac{G''}{G}\right), \tag{2}$$

$$0 \overset{!}{=} \operatorname{ric}\left(\frac{\partial}{\partial\varphi},\frac{\partial}{\partial\varphi}\right) = -\sin^2\vartheta\left(\frac{F'}{F}GG' + GG'' - 1 + (G')^2\right), \tag{3}$$

$$0 \overset{!}{=} \operatorname{ric}\left(\frac{\partial}{\partial\vartheta},\frac{\partial}{\partial\vartheta}\right) = -\left(\frac{F'}{F}GG' + GG'' - 1 + (G')^2\right). \tag{4}$$

Multiplying (2) with $-F^2G$ and (1) with $-G$ and adding the two resulting equations yields

$$0 = FGF'' + 2F^2G'' - FGF'' - 2F'G'F = 2F(FG'' - F'G')$$

and hence

$$\left(\frac{G'}{F}\right)' = \frac{G''F - G'F'}{F^2} = 0.$$

This means that $\frac{G'}{F} =: a$ is constant and non-zero, for otherwise $G' \equiv 0$ and also $G'' = 0$, a contradiction to (4). It follows that G is strictly monotonic and we can make the parameter transformation

$$r := G(\tilde{r}).$$

Then $\frac{dr}{d\tilde{r}} = G'$ and hence $dr = G' d\tilde{r}$. Abbreviating $dx^2 := dx \otimes dx$ we get

$$dr^2 = (G')^2 d\tilde{r}^2 = a^2 F^2 d\tilde{r}^2.$$

After this parameter substitution the metric takes the form

$$\begin{aligned} g &= -F(r)^2 dt^2 + \frac{1}{a^2 F(r)^2} dr^2 + r^2(\sin(\vartheta)^2 d\varphi^2 + d\vartheta^2) \\ &= -F(r)^2 dt^2 + \frac{1}{a^2 F(r)^2} dr^2 + r^2 g_{S^2} \end{aligned}$$

with a new function F, yet to be determined.
We make the following *physical assumption*: Far from our astronomical object, the spacetime should look approximately like Minkowski space

$$g_{\mathrm{Mink}} = -dt^2 + dr^2 + r^2 g_{S^2},$$

reflecting the fact that far away from the central mass special relativity is a reasonable approximation. More precisely, this means $\lim\limits_{r\to\infty} F(r) = 1$ and $a^2 = 1$. Hence the metric must have the form

$$g = -F(r)^2 dt^2 + \frac{1}{F(r)^2} dr^2 + r^2(\sin(\vartheta)^2 d\varphi^2 + d\vartheta^2).$$

For the Ricci curvature we now obtain

$$\begin{aligned} 0 &= \mathrm{ric}\left(\frac{\partial}{\partial t}, \frac{\partial}{\partial t}\right) = F^2\left((F')^2 + FF'' + 2\frac{FF'}{r}\right), \\ 0 &= \mathrm{ric}\left(\frac{\partial}{\partial \tilde{r}}, \frac{\partial}{\partial \tilde{r}}\right) = -\left(\left(\frac{F'}{F}\right)^2 + \frac{F''}{F} + 2\frac{F'}{rF}\right), \\ 0 &= \mathrm{ric}\left(\frac{\partial}{\partial \varphi}, \frac{\partial}{\partial \varphi}\right) = -\sin^2\vartheta\left(2FF'r - 1 + F^2\right), \\ 0 &= \mathrm{ric}\left(\frac{\partial}{\partial \vartheta}, \frac{\partial}{\partial \vartheta}\right) = -2FF'r + 1 - F^2. \end{aligned} \tag{5}$$

Hence

$$(rF^2)'' = (F^2 + 2rFF')' = 2(2FF' + r(F')^2 + rFF'') = 0$$

by (5). Thus rF^2 is of the form $rF^2 = br - 2m$ with constants $b, m \in \mathbb{R}$. In other words,

$$F^2 = b - \frac{2m}{r}.$$

Taking the limit shows

$$1 = \lim_{r\to\infty} F(r)^2 = b$$

Hence $F^2 = 1 - 2m/r$ and we find

$$g = -\Big(1 - \frac{2m}{r}\Big)dt^2 + \frac{1}{1 - \frac{2m}{r}}dr^2 + r^2 g_{S^2}.$$

Then we indeed have $\mathrm{ric} \equiv 0$. Note that we need to impose $r \neq 2m$.

Definition 4.1. For any $m \geq 0$ the manifold $\mathbb{R} \times ((0, 2m) \cup (2m, \infty)) \times S^2$ with the metric

$$g = -\Big(1 - \frac{2m}{r}\Big)dt^2 + \frac{1}{1 - \frac{2m}{r}}dr^2 + r^2 g_{S^2}$$

is called a *Schwarzschild spacetime*.

Summarizing the properties of Schwarzschild spacetime in physical and mathematical language we have:

physical formulation	*mathematical formulation*
radially symmetric	the standard O(3)-action on S^2, extended trivially to the r- and t-axis, is by isometries of the Schwarzschild metric
static	$\mathbb{R}$ acts isometrically by translation on the t-axis
vacuum solution	$\mathrm{ric} \equiv 0$
asymptotic to Minkowski spacetime	$g - g_{\mathrm{Mink}} \xrightarrow{r\to\infty} 0$

Definition 4.2. A curve

$$\gamma : s \mapsto (t(s), r(s), \varphi(s), \vartheta(s))$$

is called a *Schwarzschild observer*, if $r \equiv r_0$, $\varphi \equiv \varphi_0$, $\vartheta \equiv \vartheta_0$, and if γ is future directed and parametrized by proper time, i.e., $t' > 0$, and $g(\gamma', \gamma') = -1$.

Note

$$\nabla_{\frac{\partial}{\partial t}} \frac{\partial}{\partial t} = \frac{(r-2m)m}{r^3} \frac{\partial}{\partial r}.$$

For any Schwarzschild observer (with $\gamma' = t' \frac{\partial}{\partial t}$) we have

$$-1 = g(\gamma', \gamma') = (t')^2 g(\frac{\partial}{\partial t}, \frac{\partial}{\partial t}) = -(t')^2(1 - \frac{2m}{r_0}).$$

Hence

$$t' = \frac{1}{\sqrt{1 - \frac{2m}{r_0}}}$$

and therefore

$$\gamma(s) = \Big(t_0 + \frac{s}{\sqrt{1-\frac{2m}{r_0}}}, r_0, \varphi_0, \vartheta_0\Big).$$

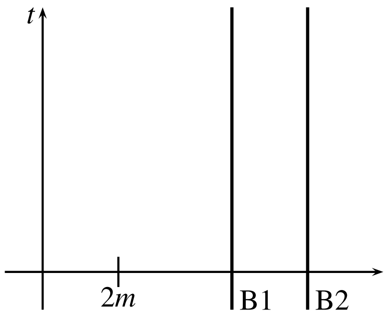


Figure 57.. Schwarzschild observer

This parametrization by proper time shows that for a Schwarzschild observer B1 with small $r_0 > 2m$, less time elapses to traverse the same cosmic time interval (measured in the coordinate t) than for a distant Schwarzschild observer with big r_0. Hence clocks run slower when under the influence of gravitation. The *Global Positioning System (GPS)* was the first technical installation where this effect had to be taken into account.
A Schwarzschild observer is subject to the acceleration

$$\frac{\nabla}{ds}\gamma' = \nabla_{\left(\frac{1}{\sqrt{1-\frac{2m}{r_0}}}\frac{\partial}{\partial t}\right)}\left(\frac{1}{\sqrt{1-\frac{2m}{r_0}}}\frac{\partial}{\partial t}\right) = \frac{1}{1-\frac{2m}{r_0}} \frac{(r_0-2m)m}{{r_0}^3}\frac{\partial}{\partial r} = \frac{m}{{r_0}^2}\frac{\partial}{\partial r}.$$

This acceleration compensates for the gravitational attraction by the central mass. It has the absolute value

$$\frac{m}{{r_0}^2}\frac{1}{\sqrt{1-\frac{2m}{r_0}}} \overset{r_0\to\infty}{\sim} \frac{m}{{r_0}^2},$$

which approximates that of a central star of mass m in Newtonian gravity, see Section 2.1. Hence m is interpreted as the mass of the astronomical object.

Definition 4.3. Let M be a semi-Riemannian manifold and let $\Phi : (-\varepsilon, \varepsilon) \to \mathrm{Isom}(M)$ be such that $\Phi(0) = \mathrm{id}_M$ and assume that $(-\varepsilon, \varepsilon) \times M \to M$ defined by $(s, p) \mapsto \Phi(s)(p)$ is smooth. Then the vector field ξ, defined by

$$\xi|_p := \frac{d}{ds}\Phi(s)(p)|_{s=0},$$

is called a *Killing vector field*.

Example 4.4. Let $M = S^2$ and

$$\Phi(s) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(s) & -\sin(s) \\ 0 & \sin(s) & \cos(s) \end{pmatrix}.$$

The corresponding Killing vector field is $\frac{\partial}{\partial \varphi}$ in polar coordinates.

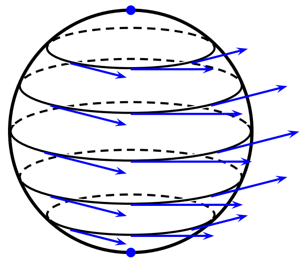


Figure 58.. Killing vector field on S^2

Lemma 4.5. *Let (M, g) be a semi-Riemannian manifold, let ξ be a Killing vector field on M and let γ be a geodesic in M. Then the function*

$$t \mapsto g(\gamma'(t), \xi|_{\gamma(t)})$$

is constant.

Proof. a) We first check that Killing vector fields have a skew symmetric covariant differential, i.e. ξ satisfies

$$\langle \nabla_X \xi, X \rangle = 0$$

for all tangent vectors X.
Indeed, since each $\Phi(s)$ is an isometry we have $\Phi(s)^* g = g$ and we get for the Lie derivative:

$$\mathscr{L}_\xi g = \frac{d}{ds}|_{s=0} \Phi(s)^* g = 0.$$

Now if X is an arbitrary vector field we find

$$\begin{aligned} 0 &= (\mathscr{L}_\xi g)(X, X) = \mathscr{L}_\xi (g(X, X)) - g(\mathscr{L}_\xi X, X) - g(X, \mathscr{L}_\xi X) \\ &= \partial_\xi (g(X, X)) - 2g([\xi, X], X) = 2g(\nabla_\xi X, X) - 2g([\xi, X], X) = 2g(\nabla_X \xi, X). \end{aligned}$$

b) Now we compute

$$\frac{d}{dt} \langle \gamma'(t), \xi(\gamma(t)) \rangle = \langle \frac{\nabla}{dt} \gamma'(t), \xi(\gamma(t)) \rangle + \langle \gamma'(t), \nabla_{\gamma'(t)} \xi(\gamma(t)) \rangle = 0 + 0 = 0. \qquad \square$$

This lemma is a version of Noether's theorem; infinitesimal symmetries (Killing vector fields) give rise to conservation laws.
In the Schwarzschild model M, $\frac{\partial}{\partial t}$ is a Killing vector field because M is static and $\frac{\partial}{\partial \varphi}$ is a Killing vector field because M is radially symmetric. Lemma 4.5 implies that for geodesics

$$\gamma(s) = (t(s), r(s), \varphi(s), \vartheta(s))$$

with $\vartheta \equiv \frac{\pi}{2}$ (i.e. in particular $\gamma' = t'\frac{\partial}{\partial t} + r'\frac{\partial}{\partial r} + \varphi'\frac{\partial}{\partial\varphi}$),

$$\text{the } \textit{energy } E := \left\langle \gamma', \frac{\partial}{\partial t} \right\rangle = -t'h \text{ and the } \textit{angular momentum } L := \left\langle \gamma', \frac{\partial}{\partial\varphi} \right\rangle = \varphi' r^2$$

are constant. Here, $h(r) := 1 - 2m/r$. With this notation, the Schwarzschild metric takes the form $g = -h(r)dt^2 + \frac{1}{h(r)}dr^2 + r^2 g_{S^2}$. For a lightlike geodesic we now have

$$0 = \langle \gamma', \gamma' \rangle = -(t')^2 \cdot h + \frac{(r')^2}{h} + r^2(\varphi')^2 = t' \cdot E + \frac{(r')^2}{h} + \varphi' \cdot L.$$

Multiplying by h, this implies the *energy equation for light particles*

$$E^2 = (r')^2 + \varphi' \cdot Lh = (r')^2 + \frac{L^2}{r^2}h.$$

For massive particles, we obtain

$$-1 = \langle \gamma', \gamma' \rangle = t'E + \frac{(r')^2}{h} + \varphi' L.$$

This yields the *energy equation for massive particles*

$$E^2 = (r')^2 + \varphi' \cdot Lh + h = (r')^2 + \left(\frac{L^2}{r^2} + 1\right)h.$$

4.1.1. Trajectories of massless particles

If $L = 0$ then φ and r' are constant and we have a simple collision orbit. So let us focus on the case $L \neq 0$. Put

$$V(r) := \frac{L^2}{r^2}h(r) = \frac{L^2}{r^2}\left(1 - \frac{2m}{r}\right).$$

We have $V(2m) = 0$, $\lim_{r\to\infty} V(r) = 0$ and $\lim_{r\to 0} V(r) = -\infty$. We determine the extrema:

$$\begin{aligned} 0 &\overset{!}{=} V'(r) \\ &= -2\frac{L^2}{r^3}\left(1 - \frac{2m}{r}\right) + \frac{L^2}{r^2}\frac{2m}{r^2} \\ &= \frac{L^2}{r^4}(-2r + 4m + 2m) \\ &= \frac{2L^2}{r^4}(-r + 3m) \end{aligned}$$

Thus there is only one extremum at $r = 3m$. Because of the behavior of V for $r \searrow 0$ and $r \to \infty$, $r = 3m$ must be a maximum with $V(3m) = \frac{L^2}{27m^2}$.

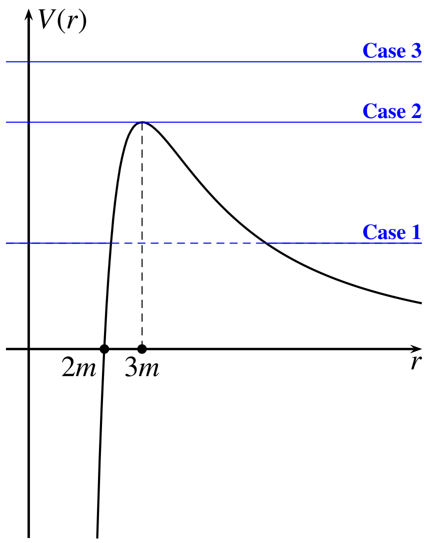


Figure 59.. Energy diagram for massless particles in Schwarzschild

The energy equation takes the form $E^2 = (r')^2 + V(r)$. In particular, $V(r) \leq E^2$.

Case 1: $E^2 < \dfrac{L^2}{27m^2}$.

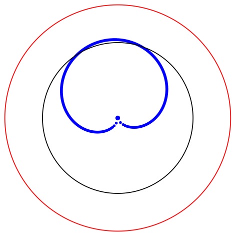


Figure 60.. Collision-collision orbit

(a) $r_0 < 3m$: collision-collision orbit.
(b) $r_0 > 3m$: fly-by orbit. Unlike in Newtonian gravity, even light is deflected under the influ-

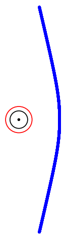


Figure 61.. Fly-by orbit

ence of gravitation. Indeed,

$$\lim_{s\to\infty}\varphi(s) - \lim_{s\to-\infty}\varphi(s) = \int_{-\infty}^{\infty}\varphi'(s)\,ds = \int_{-\infty}^{\infty}\frac{L}{r(s)^2}ds \neq 0$$

if $L \neq 0$. Comparing the apparent distance of stars with the one observed during an eclipse one can measure this light deflection. This has been the first experimental confirmation of general relativity.

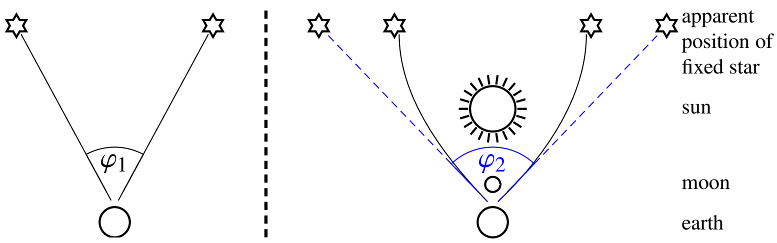


Figure 62.. Deflection of light

Case 2: $E^2 = \dfrac{L^2}{27m^2}$.
(a) $r \equiv 3m$: exceptional orbit, the *photon sphere*.
(b) $r_0 < 3m$ or $r_0 > 3m$: spiral orbits.

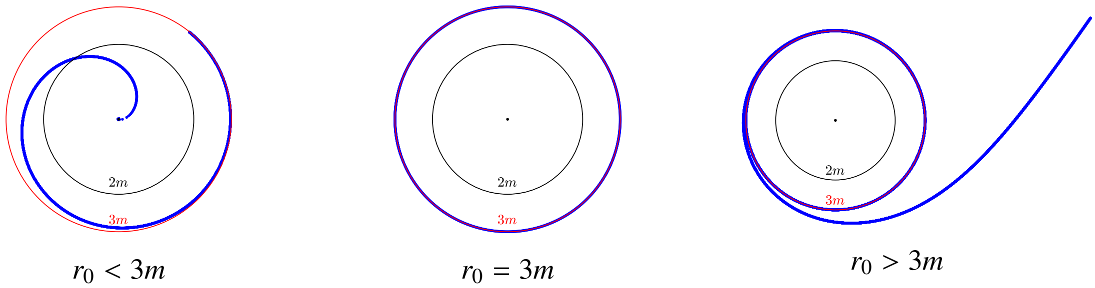


Figure 63.. Spiral orbits and photon sphere

Case 3: $E^2 > \frac{L^2}{27m^2}$: collision-escape orbit.

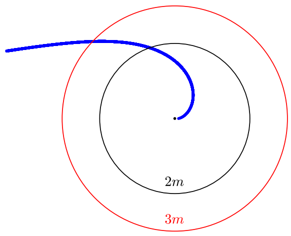


Figure 64.. Collision-escape orbit

These orbits not being straight lines has an impact on the sight angle. Astronomical objects appear to be bigger than they are.

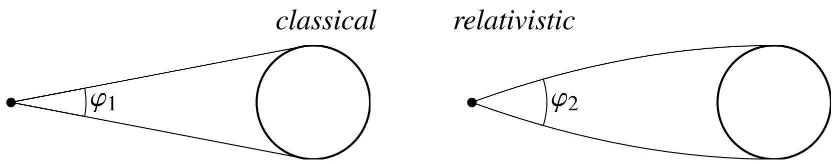


Figure 65.. Classical versus relativistic sight angle

4.1.2. Orbits of massive particles

Now set

$$V(r) := \left(\frac{L^2}{r^2} + 1\right) h(r).$$

We have $V(2m) = 0$, $\lim\limits_{r\to\infty} V(r) = 1$ and $\lim\limits_{r\to 0} V(r) = -\infty$. The local extrema are at

$$r_{1,2} = \frac{L^2}{2m} \pm L\sqrt{\frac{L^2}{4m^2} - 3}.$$

Case 1: $L^2 < 12m^2$, i.e. there are no local extrema.

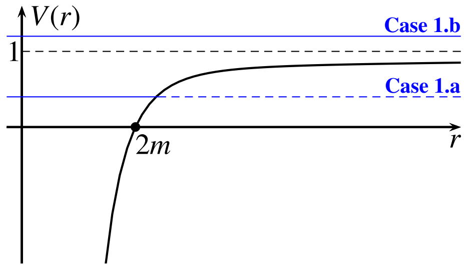


Figure 66.. Energy diagram for massive particles in Schwarzschild for $L^2 < 12m^2$

(a) $E^2 < 1$: collision-collision orbit.

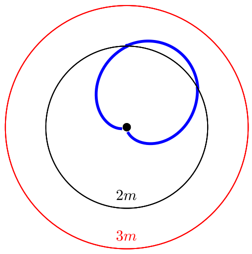


Figure 67.. Collision-collision orbit

(b) $E^2 \geq 1$: collision-escape orbit.

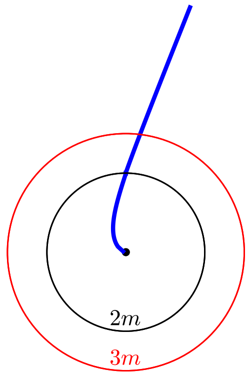


Figure 68.. Collision-escape orbit

Case 2: $12m^2 \leq L^2 < 16m^2$.

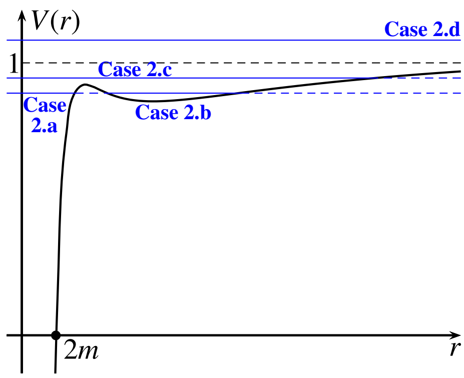


Figure 69.. Energy diagram for massive particles in Schwarzschild for $12m^2 \leq L^2 < 16m^2$

(a) $E^2 < V(r_1)$ and $r_0 < r_1$: collision-collision orbit.

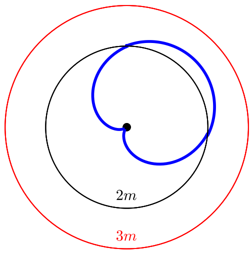


Figure 70.. Collision-collision orbit

(b) $E^2 < V(r_1)$ and $r_0 > r_1$: bounded orbit.

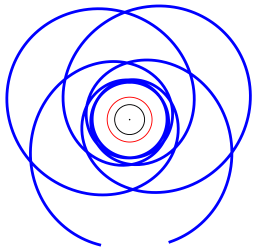


Figure 71.. Bounded orbit

(c) $V(r_1) < E^2 < 1$: collision-collision orbit.

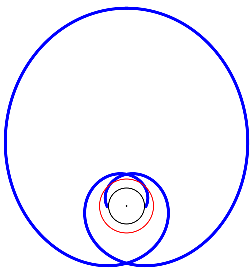


Figure 72.. Collision-collision orbit

(d) $E^2 \geq 1$: collision-escape orbit.

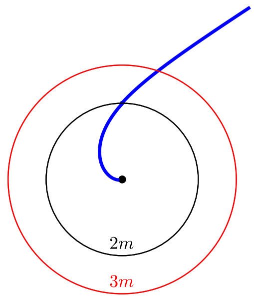


Figure 73.. Collision-escape orbit

Case 3: $L^2 > 16m^2$.

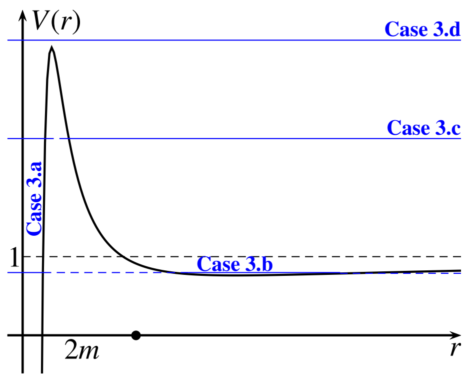


Figure 74.. Energy diagram for massive particles in Schwarzschild for $L^2 < 16m^2$

(a) $E^2 < V(r_1)$ and $r_0 < r_1$: collision-collision orbit.

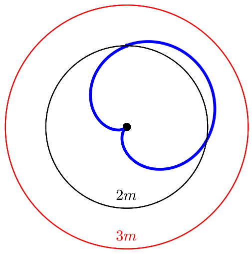


Figure 75.. Collision-collision orbit

(b) $V(r_2) < E^2 < 1$ and $r_0 > r_1$: bounded orbit.

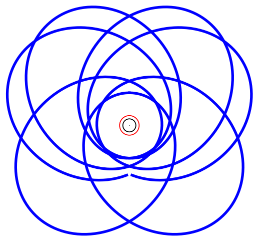


Figure 76.. Bounded orbit

(c) $1 \leq E^2 < V(r_1)$ and $r_0 > r_1$: fly-by orbit.

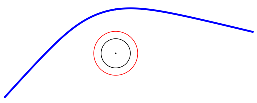


Figure 77.. Fly-by orbit

(d) $E^2 > V(r_1)$: collision-escape orbit.

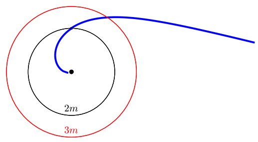


Figure 78.. Collision-escape orbit

Definition 4.6. The constant $2m$ is called the *Schwarzschild radius*.

4.1.3. Kruskal coordinates

We note that the Schwarzschild metric is ill defined at $r = 2m$ and at $r = 0$. At $r = 2m$ the function $h(r)$ vanishes and at $r = 0$ the coefficient r^2 in the metric vanishes. One may wonder whether this is due to an actual singular behavior of the metric as r approaches $2m$ or 0, respectively, or if the singularity is only caused by a suboptimal choice of coordinates.
In fact, for $m = 0$, when both singular values coincide, coordinates can be changed in such a way that the singularity at $r = 0$ disappears. Namely, for $m = 0$ the Schwarzschild metric is noting but the Minkowski metric where the Euclidean part on $\mathbb{R}^3$ is parametrized by polar coordinates. Switching to Cartesian coordinates will make the singularity disappear.
It is indeed possible to remove the singularity at $r = 2m$ for $m > 0$ in a similar fashion. Set

$$f(r) := (r - 2m)e^{\frac{r}{2m}-1}.$$

Then $f : (0, \infty) \to (-\frac{2m}{e}, \infty)$ is a diffeomorphism, because

$$f'(r) = e^{\frac{r}{2m}-1} + (r - 2m)\frac{1}{2m}e^{\frac{r}{2m}-1} = \frac{r}{2m}e^{\frac{r}{2m}-1} > 0.$$

We have $f((0, 2m)) = (-\frac{2m}{e}, 0)$ and $f((2m, \infty)) = (0, \infty)$.

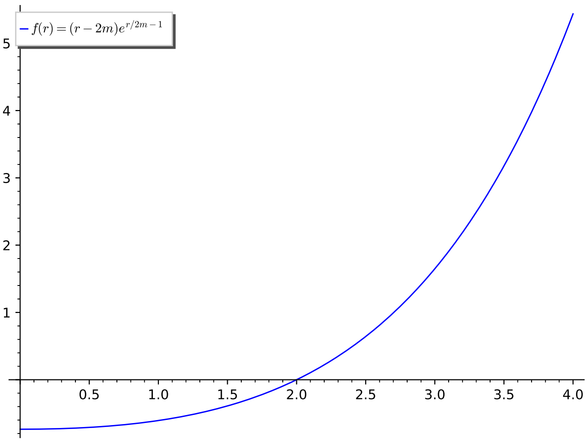


Figure 79.. Auxiliary function for coordinate change

We consider the smooth map

$$\{(u,v)\,|\,uv > -\tfrac{2m}{e},\ u \neq 0,\ v \neq 0\} \to \{(t,r)\,|\,r > 0,\ r \neq 2m\}$$

given by

$$r = f^{-1}(uv), \qquad t = 2m \cdot \ln\left(\left|\frac{v}{u}\right|\right).$$

The map is a local diffeomorphism because

$$\det\begin{pmatrix} \frac{\partial r}{\partial u} & \frac{\partial r}{\partial v} \\ \frac{\partial t}{\partial u} & \frac{\partial t}{\partial v} \end{pmatrix} = \det\begin{pmatrix} v(f^{-1})'(uv) & u(f^{-1})'(uv) \\ -\frac{2m}{u} & \frac{2m}{v} \end{pmatrix} = 4m(f^{-1})'(uv) > 0.$$

It is surjective but not injective since (u,v) and $(-u,-v)$ are mapped to the same point.

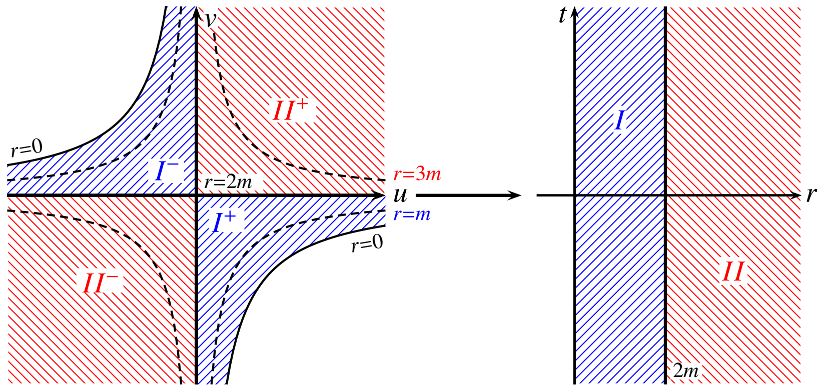


Figure 80.. Kruskal coordinates

We take the product with the identity on S^2 and pull the Schwarzschild metric back. We compute

$$
\begin{aligned}
dt &= 2m\, d(\ln|v| - \ln|u|) = 2m\Big(\frac{dv}{v} - \frac{du}{u}\Big),\\
dt\otimes dt &= 4m^2\Big(\frac{du\otimes du}{u^2} - \frac{du\otimes dv}{uv} - \frac{dv\otimes du}{uv} + \frac{dv\otimes dv}{v^2}\Big),\\
dr &= d(f^{-1}(uv)) = (f^{-1})'(uv)(udv + vdu) = \frac{udv + vdu}{f'(r)} = \frac{2m}{r}e^{1-\frac{r}{2m}}(udv + vdu),\\
dr\otimes dr &= \frac{4m^2}{r^2}e^{2-\frac{r}{m}}(v^2du\otimes du + uvdu\otimes dv + uvdv\otimes du + u^2dv\otimes dv).
\end{aligned}
$$

Using $uv = f(r) = (r-2m)e^{\frac{r}{2m}-1}$ we find

$$
\begin{aligned}
-h(r)dt\otimes dt &= -\frac{4m^2h(r)}{f(r)^2}(v^2du\otimes du - uvdu\otimes dv - uvdv\otimes du + u^2dv\otimes dv)\\
&= \frac{4m^2}{r(r-2m)}e^{2-\frac{r}{m}}(-v^2du\otimes du + uvdu\otimes dv + uvdv\otimes du - u^2dv\otimes dv),\\
\frac{1}{h(r)}dr\otimes dr &= \frac{4m^2}{r(r-2m)}e^{2-\frac{r}{m}}(v^2du\otimes du + uvdu\otimes dv + uvdv\otimes du + u^2dv\otimes dv),
\end{aligned}
$$

and therefore

$$
\begin{aligned}
-h(r)dt\otimes dt + \frac{1}{h(r)}dr\otimes dr &= \frac{4m^2}{r(r-2m)}e^{2-\frac{r}{m}}\cdot 2uv(du\otimes dv + dv\otimes du)\\
&= \frac{8m^2}{r}e^{1-\frac{r}{2m}}(du\otimes dv + dv\otimes du).
\end{aligned}
$$

This gives us for the Schwarzschild metric

$$
\begin{aligned}
g &= \tfrac{8m^2}{r}e^{1-\frac{r}{2m}}(du\otimes dv + dv\otimes du) + r^2g_{S^2}\\
&= \tfrac{8m^2}{f^{-1}(uv)}e^{1-\frac{f^{-1}(uv)}{2m}}(du\otimes dv + dv\otimes du) + f^{-1}(uv)^2g_{S^2}.
\end{aligned}
$$

The crucial observation is now that this metric extends smoothly to $u = 0$ and $v = 0$. By passing to the *Kruskal coordinates* u and v we can extend the Schwarzschild metric smoothly across the event horizon $r = 2m$.
In other words, if we define the *Kruskal plane*

$$
\mathsf{Kr} = \{(u, v)\,|\, uv > -\tfrac{2m}{e}\}
$$

equipped with the metric $g_{\mathsf{Kr}} = \frac{8m^2}{r}e^{1-\frac{r}{2m}}(du\otimes dv + dv\otimes du)$ where $f(r) = uv$, then the manifold $\mathsf{Kr}\times S^2$ with $(\mathbb{R}\times\{0\}\cup\{0\}\times\mathbb{R})\times S^2$ removed and equipped with the metric $g_{\mathsf{Kr}} + r^2g_{S^2}$ maps as a local isometry onto the Schwarzschild spacetime. The regions $I^\pm$ map onto the interior region I and the regions $II^\pm$ map onto the outer region II. Note that Kr and hence $\mathsf{Kr}\times S^2$ is connected. On the outer region II, the time orientation of the Schwarzschild spacetime is characterized by $t' > 0$. In Kruskal coordinates this means on II^+:

$$
0 < t' = 2m\Big(\frac{v'}{v} - \frac{u'}{u}\Big) = 2m\frac{v'u - u'v}{uv} \quad\Longleftrightarrow\quad v'u - u'v > 0.
$$

This shows that $-\frac{\partial}{\partial u}$ and $\frac{\partial}{\partial v}$ are future-directed lightlike vector fields on II^+. We can now extend this time-orientation in a unique manner continuously to all of $\mathsf{Kr} \times S^2$ by demanding that $-\frac{\partial}{\partial u}$ and $\frac{\partial}{\partial v}$ are future-directed everywhere. Future-directed curves in Kr move upwards or to the left or both simultaneously. Thus the cannot leave the region I^-. The union of II^+ and I^- yields a model for a *black hole*. Anything that has passed the event horizon from the outer region can never pass it again. Similarly, The union of II^+ and I^+ yields a model for a *white hole*. Nothing can pass the event horizon from the outer region to the interior region, only passages in the opposite direction are possible.

4.1.4. The singularity at $r = 0$

Now, how about the singularity at $r = 0$? Can we find a suitable coordinate change which will remove this singularity? We know it is possible if $m = 0$ but how about $m > 0$?
It is impossible indeed. This can be seen by considering a curvature quantity which is defined inpendently of the choice of a coordinate system and which explodes as $r \to 0$. The scalar curvature is unfortunately not helpful because it vanishes identically since the Schwarzschild spacetime is Ricci-flat. There is another curvature function that helps us out here, so *Kretschmann scalar curvature*. It is defined as the square of the full Riemann curvature tensor with respect to the metric induced on the space of curvature tensors by the Lorentzian metric. This is defined invariantly, but can be expressed in coordinates as

$$Km = g(R, R) = \sum g_{ii'} g^{jj'} g^{kk'} g^{ll'} R^i_{jkl} R^{i'}_{j'k'l'}$$

where we sum over all indices. In the case of the Schwarzschild spacetime one finds

$$Km = \frac{48m^2}{r^6}.$$

In particular, $Km \to \infty$ as $r \to 0$. Thus Schwarzschild cannot be extended to $r = 0$. Note that this argument fails if $m = 0$.
At $r = 0$ we have a true singularity of the spacetime. The Penrose singularity theorem says that under certain natural conditions such singularities must necessarily form, see e.g. [7, Ch. 14].

4.2. Rotating black holes - the Kerr solution

Our goal in this section will be to generalize the Schwarzschild solution and allow for rotation of the gravitating object. This solution was found by Roy Patrick Kerr in 1963, a mathematician born in 1934 in New Zealand.

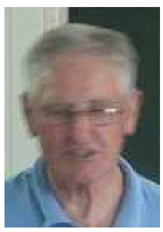


Figure 81.. Roy Patrick Kerr (*1934)

4.2.1. The ansatz

The ansatz is the following: On a suitable open subset of $\mathbb{R}^2 \times S^2$ we will define the *Kerr metric* $g_{(m,a)}$ depending on 2 fixed parameters $m > 0$ and $a \in \mathbb{R}$. The physical interpretation of m will be the mass of the black hole just like for the Schwarzschild solution and a will be interpreted as angular momentum per unit mass.

On the first factor of the product manifold $\mathbb{R}^2 \times S^2$ we use standard coordinates which we denote by (t, r). On the sphere $S^2 \subset \mathbb{R}^3$ we consider the usual polar coordinate functions

$$\varphi : S^2 \setminus \{(0,0,\pm 1)\} \to S^1 = \mathbb{R}/2\pi\mathbb{Z},$$
$$\vartheta : S^2 \setminus \{(0,0,\pm 1)\} \to (0,\pi),$$

in such a way that $S^2 \setminus \{(0,0,\pm 1)\}$ is parametrized by

$$x = \sin\vartheta\cos\varphi, \quad y = \sin\vartheta\sin\varphi, \quad z = \cos\vartheta.$$

Next we define the two auxiliary functions $\rho, \Delta : \mathbb{R}^2 \times S^2 \to \mathbb{R}$ by

$$\rho^2 := r^2 + a^2\cos^2\vartheta,$$
$$\Delta := r^2 - 2mr + a^2.$$

We consider the Riemannian metric $g_{(m,a)}$ whose components in the basis $(\partial_t, \partial_r, \partial_\vartheta, \partial_\varphi)$ are given by

$$\begin{pmatrix} -1 + \frac{2mr}{\rho^2} & 0 & 0 & -\frac{2mra\sin^2\vartheta}{\rho^2} \\ 0 & \frac{\rho^2}{\Delta} & 0 & 0 \\ 0 & 0 & \rho^2 & 0 \\ -\frac{2mra\sin^2\vartheta}{\rho^2} & 0 & 0 & (r^2 + a^2 + \frac{2mra^2\sin^2\vartheta}{\rho^2})\sin^2\vartheta \end{pmatrix}. \tag{6}$$

Remark 4.7. **(1)** In the literature these expressions are called the components of the Kerr metric in *Boyer-Lindquist coordinates*.
(2) If we set $a = 0$ we obtain the Schwarzschild metric defined on $\mathbb{R} \times ((0,2m) \cup (2m,\infty)) \times S^2$.
(3) If we set $m = 0$ we obtain the Minkowski metric in coordinates t, r, ϑ, φ on $\mathbb{R} \times U$ where U is an open subset of $\mathbb{R}^3$ parametrized by

$$x = \sqrt{r^2 + a^2}\sin\vartheta\cos\varphi, \quad y = \sqrt{r^2 + a^2}\sin\vartheta\sin\varphi, \quad z = r\cos\vartheta,$$

see the check B.2.1 by SageMath.
(4) Note that the coordinate r may take values in all of $\mathbb{R}$ including negative ones.

From now on we assume $m > 0$ and $a > 0$ unless stated otherwise. The first question that arises is: On which subset of $\mathbb{R}^2 \times S^2$ does $g_{(m,a)}$ give a well-defined Lorentzian metric?
Looking at the form of $g_{(m,a)}$ in (6) we see that we have to impose the following conditions:
(a) We need $\rho \neq 0$. Therefore we exclude the *ring singularity*

$$\Sigma := \rho^{-1}(0) = \mathbb{R} \times \{0\} \times \{\sigma \in S^2 \,|\, \vartheta(\sigma) = \tfrac{\pi}{2}\}.$$

(b) We need $\Delta \neq 0$. Therefore we exclude $H := \Delta^{-1}(0)$. In order to determine the solutions r to the equation

$$0 = \Delta = r^2 - 2mr + a^2$$

we distinguish three cases:

1. If $a > m$ the equation has no real solution r and consequently $H = \emptyset$. We call this metric a *rapidly rotating black hole*.
2. If $a = m$ the only solution is $r = m$ and we have $H = \mathbb{R} \times \{m\} \times S^2$. We call this metric an *extremal black hole*.
3. If $a < m$ we have two real solutions $r_\pm := m \pm \sqrt{m^2 - a^2}$ and we put

$$H_\pm := \mathbb{R} \times \{r_\pm\} \times S^2, \quad H = H_- \cup H_+.$$

We call H_- and H_+ the *inner* and the *outer horizon* respectively and we call the metric a *slowly rotating black hole*.

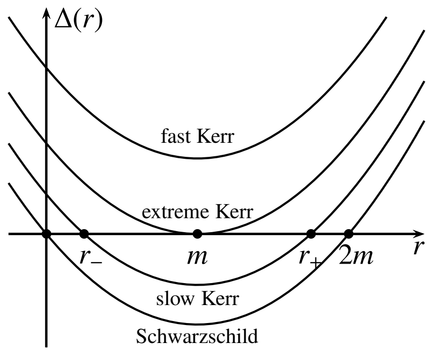


Figure 82.. The function Δ

(c) We need $\sin\vartheta \neq 0$. Therefore we exclude $A := A_+ \cup A_-$ where

$$A_\pm := \mathbb{R}^2 \times \{(0, 0, \pm 1)\}.$$

The following picture shows a slice $t =$ const in $\mathbb{R}^2 \times S^2$ in the case of a slowly rotating black hole. The labels $\Sigma, H_\pm, A_\pm$ denote the slices of the respective subsets. Now we have removed all subsets of $\mathbb{R}^2 \times S^2$ where the definition of $g_{(m,a)}$ encounters problems. In other words, $g_{(m,a)}$ is a well-defined smooth $(0, 2)$-tensor field on $\mathbb{R}^2 \times S^2 \setminus (\Sigma \cup H \cup A)$. Next we have to check that the tensor field $g_{(m,a)}$ is actually a Lorentzian metric on $\mathbb{R}^2 \times S^2 \setminus (\Sigma \cup H \cup A)$.
We first need the following lemma.

Lemma 4.8 (Boyer-Lindquist identities).

$$g_{\varphi\varphi} + a\sin^2\vartheta g_{\varphi t} = (r^2 + a^2)\sin^2\vartheta, \tag{BL1}$$
$$g_{t\varphi} + a\sin^2\vartheta g_{tt} = -a\sin^2\vartheta, \tag{BL2}$$
$$a g_{\varphi\varphi} + (r^2 + a^2) g_{t\varphi} = \Delta a \sin^2\vartheta, \tag{BL3}$$
$$a g_{t\varphi} + (r^2 + a^2) g_{tt} = -\Delta. \tag{BL4}$$

Proof. This is a straighforward computation, see B.2.2. □

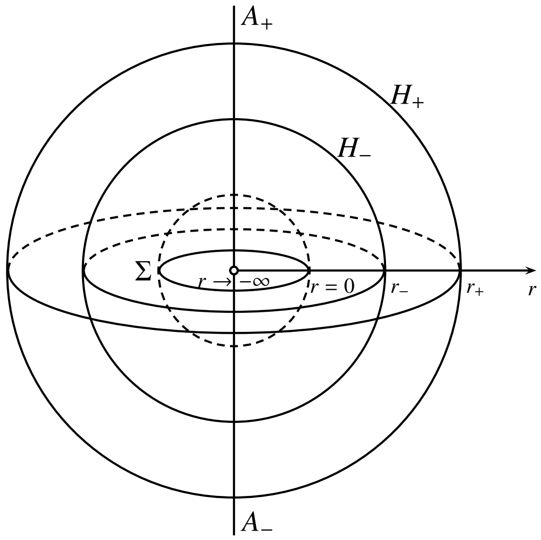


Figure 83.. Ring singularity, horizons, and axis

Reordering the basis vectors to $(\partial_t, \partial_\varphi, \partial_r, \partial_\vartheta)$ the tensor field $g_{(m,a)}$ has the matrix representation

$$\begin{pmatrix} -1 + \frac{2mr}{\rho^2} & -\frac{2mra\sin^2\vartheta}{\rho^2} & 0 & 0 \\ -\frac{2mra\sin^2\vartheta}{\rho^2} & (r^2 + a^2 + \frac{2mra^2\sin^2\vartheta}{\rho^2})\sin^2\vartheta & 0 & 0 \\ 0 & 0 & \frac{\rho^2}{\Delta} & 0 \\ 0 & 0 & 0 & \rho^2 \end{pmatrix}.$$

Lemma 4.9. *(1)* $g_{tt}g_{\varphi\varphi} - g_{t\varphi}^2 = -\Delta\sin^2\vartheta$,
(2) $\det(g_{ij}) = -\rho^4\sin^2\vartheta$.

Proof. Using (BL1), (BL2), and (BL4) we compute

$$\begin{aligned} g_{tt}g_{\varphi\varphi} - g_{t\varphi}^2 &= g_{tt}((r^2 + a^2)\sin^2\theta - a\sin^2\theta g_{\varphi t}) - g_{t\varphi}(-a\sin^2\theta - a\sin^2\theta g_{tt}) \\ &= (r^2 + a^2)\sin^2\theta g_{tt} + a\sin^2\theta g_{t\varphi} \\ &= -\Delta\sin^2\theta. \end{aligned}$$

This shows the first identity and the second one follows immediately. □

We distinguish two cases:

(1) $\Delta > 0$: The component matrix is

$$(g_{ij}) = \begin{pmatrix} \widetilde{g}_{ij} & & \\ & + & \\ & & + \end{pmatrix}$$

where $+$ denotes a positive entry and $\widetilde{g} \in \mathbb{R}^{2\times 2}$ has $\det(\widetilde{g}) < 0$. Thus (g_{ij}) has Lorentzian signature.

(2) $\Delta < 0$: The component matrix is

$$(g_{ij}) = \begin{pmatrix} \widetilde{g}_{ij} & & \\ & - & \\ & & + \end{pmatrix}$$

where $\pm$ denote a positive/negative entry and $\widetilde{g} \in \mathbb{R}^{2\times 2}$ has $\det(\widetilde{g}) > 0$. In order to conclude that (g_{ij}) has Lorentzian signature it remains to show that all eigenvalues of $\widetilde{g}$ are positive. This follows if we can show that $\widetilde{g}_{11}$ is positive. Using that $\cos^2\vartheta \leq 1$ we obtain

$$\widetilde{g}_{11} = -1 + \frac{2mr}{\rho^2} = -\frac{r^2 + a^2\cos^2\vartheta - 2mr}{\rho^2} \geq -\frac{r^2 + a^2 - 2mr}{\rho^2} = -\frac{\Delta}{\rho^2} > 0.$$

Hence in both cases we obtain a Lorentzian metric.

4.2.2. Extension across the axis

We have shown that $g_{(m,a)}$ defines a Lorentzian metric on $\mathbb{R}^2 \times S^2 \setminus (\Sigma \cup H \cup A)$. Our next aim is to extend $g_{(m,a)}$ across the subset A. As a first step we recall that we have parametrized $\mathbb{R}^3$ by spherical coordinates

$$x = R\sin\vartheta\cos\varphi, \quad y = R\sin\vartheta\sin\varphi, \quad z = R\cos\vartheta.$$

It follows that

$$xdy - ydx = R^2\sin^2\vartheta d\varphi \tag{7}$$

and therefore $\sin^2\vartheta d\varphi$ extends uniquely to a smooth 1-form on all of S^2 (in contrast to $d\varphi$ itself). The extended 1-form vanishes at the points $(0, 0, \pm 1)$.

Lemma 4.10. *On $\mathbb{R}^2 \times S^2 \setminus (\Sigma \cup H \cup A)$ we have*

$$g_{(m,a)} = \frac{\rho^2}{\Delta}dr^2 + \rho^2 g_{S^2} + a^2\sin^4\vartheta d\varphi^2 - dt^2 + \frac{2mr}{\rho^2}(dt - a\sin^2\vartheta d\varphi)^2 \tag{8}$$

where g_{S^2} denotes the standard metric on S^2.

Proof. See Exercise 4.8. □

Corollary 4.11. *The metric $g_{(m,a)}$ can be extended uniquely to a Lorentzian metric on $\mathbb{R}^2 \times S^2 \setminus (\Sigma \cup H)$.*

Proof. All terms on the right hand side of (8) can be extended smoothly across A. The extension is unique since the complement of A is dense. It remains to show that the extension has Lorentzian signature on A. To see this recall that on A we have $\sin^2 \vartheta d\varphi = 0$ and thus

$$g_{(m,a)} = \frac{\rho^2}{\Delta} dr^2 + \rho^2 g_{S^2} - dt^2 + \frac{2mr}{\rho^2} dt^2 = \frac{\rho^2}{\Delta} dr^2 + \rho^2 g_{S^2} - \frac{\Delta}{\rho^2} dt^2.$$

Thus in both cases $\Delta > 0$ and $\Delta < 0$ the extension of $g_{(m,a)}$ has Lorentzian signature on A. □

Definition 4.12. The connected components of $\mathbb{R}^2 \times S^2 \setminus (\Sigma \cup H)$ are called *Boyer-Lindquist blocks*. We will denote them as follows:

slowly rotating case:	$I := \mathbb{R} \times (r_+, \infty) \times S^2$
	$II := \mathbb{R} \times (r_-, r_+) \times S^2$
	$III := \mathbb{R} \times (-\infty, r_-) \times S^2 \setminus \Sigma$
extremal case:	$I := \mathbb{R} \times (m, \infty) \times S^2$
	$III := \mathbb{R} \times (-\infty, m) \times S^2 \setminus \Sigma$
rapidly rotating case:	$I := III := \mathbb{R} \times \mathbb{R} \times S^2 \setminus \Sigma.$

4.2.3. Isometries and special submanifolds

In the following we denote by $\mathrm{Kerr}_{m,a}$ the manifold $\mathbb{R} \times \mathbb{R} \times S^2 \setminus (\Sigma \cup H)$ equipped with the metric $g_{(m,a)}$. We will sometimes denote $g_{(m,a)}$ by $\langle \cdot, \cdot \rangle$. The elements of $\mathrm{Kerr}_{m,a}$ will be denoted by (t, r, σ). We examine some of the isometries of $\mathrm{Kerr}_{m,a}$:

(1) Let $t_0 \in \mathbb{R}$ and consider the translation

$$\mathcal{T}_{t_0} : \quad \mathrm{Kerr}_{m,a} \to \mathrm{Kerr}_{m,a}, \quad (t, r, \sigma) \mapsto (t + t_0, r, \sigma).$$

We observe that

$$\mathcal{T}_{t_0}^*(d\vartheta) = d\vartheta, \quad \mathcal{T}_{t_0}^*(d\varphi) = d\varphi, \quad \mathcal{T}_{t_0}^*(dr) = dr, \quad \mathcal{T}_{t_0}^*(dt) = dt,$$

and that the coefficients of $g_{(m,a)}$ are independent of t. It follows that $\mathcal{T}_{t_0}^* g_{(m,a)} = g_{(m,a)}$, i.e. $\mathcal{T}_{t_0}$ is an isometry for every $t_0 \in \mathbb{R}$. Since for every $p \in \mathrm{Kerr}_{m,a}$ we have

$$\frac{d}{ds} \mathcal{T}_s(p)|_{s=0} = \partial_t|_p$$

the vector field ∂_t is a Killing vector field on $\mathrm{Kerr}_{m,a}$.
(2) Let $\varphi_0 \in \mathbb{R}$ and consider

$$\mathcal{R}_{\varphi_0}: \quad \mathrm{Kerr}_{m,a} \to \mathrm{Kerr}_{m,a}, \quad (t, r, \sigma) \mapsto (t, r, R_{\varphi_0}\sigma),$$

where $R_{\varphi_0}\colon S^2 \to S^2$ is the rotation

$$R_{\varphi_0} = \begin{pmatrix} \cos\varphi_0 & -\sin\varphi_0 & 0 \\ \sin\varphi_0 & \cos\varphi_0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

In polar coordinates we have $(t, r, \vartheta, \varphi) \mapsto (t, r, \vartheta, \varphi + \varphi_0)$ and as above it follows that $\mathcal{R}_{\varphi_0}$ is an isometry for every $\varphi_0 \in \mathbb{R}$. Since for every $p \in \mathrm{Kerr}_{m,a}$ we have

$$\frac{d}{ds}\mathcal{R}_s(p)|_{s=0} = \partial_\varphi|_p$$

the vector field ∂_φ is a Killing vector field on $\mathrm{Kerr}_{m,a}$. Moreover if $\varphi_0 \neq 0 \mod 2\pi$, the set of fixed points of $\mathcal{R}_{\varphi_0}$ is

$$\mathrm{Fix}(\mathcal{R}_{\varphi_0}) = \mathbb{R}^2 \times \{(0, 0, \pm 1)\} = A.$$

Thus the axis A is a 2-dimensional totally geodesic submanifold of $\mathrm{Kerr}_{m,a}$.
(3) Let

$$\varepsilon: \quad \mathrm{Kerr}_{m,a} \to \mathrm{Kerr}_{m,a}, \quad (t, r, \sigma) \mapsto (t, r, S\sigma),$$

where $S\colon S^2 \to S^2$ is the reflection

$$S = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

In polar coordinates, S is the map $(t, r, \vartheta, \varphi) \mapsto (t, r, \pi - \vartheta, \varphi)$. Since $\sin^2(\pi - \vartheta) = \sin^2\vartheta$ and $d(\pi - \vartheta) = -d\vartheta$ and $d\vartheta$ occurs only quadratic in $g_{(m,a)}$ it follows that ε is an isometry. The set of fixed points of ε is the equatorial hyperplane

$$\mathrm{Eq} := \mathrm{Fix}(\varepsilon) = \mathbb{R}^2 \times \left\{\sigma \in S^2 \,\middle|\, \vartheta(\sigma) = \frac{\pi}{2}\right\} \setminus (\Sigma \cup H).$$

Thus Eq is a 3-dimensional totally geodesic submanifold of $\mathrm{Kerr}_{m,a}$. The isometries in (2) and (3) are the ones which survive from the O(3)-symmetries of the Schwarzschild solution.
(4) Let $t_0 \in \mathbb{R}$ and let the two maps $\Phi_{t_0}, \Psi_E\colon \mathrm{Kerr}_{m,a} \to \mathrm{Kerr}_{m,a}$ be defined by

$$\Phi_{t_0}(t, r, \sigma) = (2t_0 - t, r, \sigma), \quad \Psi_E(t, r, \sigma) = (t, r, S_E\sigma),$$

where $S_E\colon S^2 \to S^2$ is the reflection about a two-dimensional plane containing both points $(0, 0, \pm 1)$. If we allow negative values of a in the definition of $g_{(m,a)}$ we obtain

$$\Phi_{t_0}^* g_{(m,a)} = \Psi_E^* g_{(m,a)} = g_{(m,-a)}.$$

Thus the composition $\Phi_{t_0} \circ \Psi_E\colon \mathrm{Kerr}_{m,a} \to \mathrm{Kerr}_{m,a}$ is an isometry and the set of fixed points

$$\mathrm{Fix}(\Phi_{t_0} \circ \Psi_E) = (\{t_0\} \times \mathbb{R} \times (S^2 \cap E)) \setminus (\Sigma \cup H)$$

is a 2-dimensional totally geodesic submanifold of $\mathrm{Kerr}_{m,a}$. Unlike in Schwarzschild spacetime, the hypersurfaces of constant t are not totally geodesic submanifolds of $\mathrm{Kerr}_{m,a}$.

We examine further interesting submanifolds of $\mathrm{Kerr}_{m,a}$.

(5) Let $r_0 \in \mathbb{R}$ and $\vartheta_0 \in (0, \pi)$. The set

$$\mathrm{KO}_{r_0,\vartheta_0} := \mathbb{R} \times \{r_0\} \times \{\sigma \in S^2 \mid \vartheta(\sigma) = \vartheta_0\} \setminus (\Sigma \cup H)$$

is called the *Killing orbit* for r_0 and ϑ_0. It is a submanifold of $\mathrm{Kerr}_{m,a}$ which is not totally geodesic. But it is interesting to note that the metric $g_{(m,a)}$ on $\mathrm{KO}_{r_0,\vartheta_0}$ has constant coefficients with respect to the basis vectors $(\partial_t, \partial_\varphi)$

$$\begin{pmatrix} -1 + \frac{2mr_0}{\rho_0^2} & -\frac{2mr_0 a \sin^2\vartheta_0}{\rho_0^2} \\ -\frac{2mr_0 a \sin^2\vartheta_0}{\rho_0^2} & (r_0^2 + a^2 + \frac{2mr_0 a^2 \sin^2\vartheta_0}{\rho_0^2}) \sin^2\vartheta_0 \end{pmatrix}$$

where $\rho_0^2 := r_0^2 + a^2 \cos^2\vartheta_0$. It follows that $\mathrm{KO}_{r_0,\vartheta_0}$ is flat.

(6) The set

$$\mathrm{Thr} := \mathbb{R} \times \{0\} \times (S^2 \setminus \Sigma)$$

is called the *throat* of $\mathrm{Kerr}_{m,a}$. The metric $g_{(m,a)}$ on Thr with respect to the basis vectors $(\partial_t, \partial_\vartheta, \partial_\varphi)$ is given by

$$\begin{pmatrix} -1 & 0 & 0 \\ 0 & a^2\cos^2\vartheta & 0 \\ 0 & 0 & a^2\sin^2\vartheta \end{pmatrix}.$$

We define a local parametrization of the disk

$$D(a) := \{(\xi^1, \xi^2) \in \mathbb{R}^2 \mid (\xi^1)^2 + (\xi^2)^2 < a\}$$

by setting

$$\xi^1 := a \sin\vartheta \cos\varphi, \quad \xi^2 := a \sin\vartheta \sin\varphi, \quad \vartheta \in \left(0, \frac{\pi}{2}\right), \ \varphi \in S^1,$$

and we compute

$$(d\xi^1)^2 + (d\xi^2)^2 = a^2\cos^2\vartheta (d\vartheta)^2 + a^2 \sin^2\vartheta (d\varphi)^2.$$

The throat has two connected components

$$\mathrm{Thr}_\pm := \mathbb{R} \times \{0\} \times \left\{\sigma \in S^2 \,\middle|\, \vartheta(\sigma) \lessgtr \frac{\pi}{2}\right\}.$$

By the above formulas, we obtain two isometries

$$\mathrm{Thr}_\pm \to (\mathbb{R} \times D(a), -dt^2 + (d\xi^1)^2 + (d\xi^2)^2).$$

Hence both connected components of the throat are isometric to an open subset of 2 + 1-dimensional Minkowski space. In particular, the throat is flat.

Definition 4.13. A Lorentzian manifold is called *stationary* if it admits a timelike Killing vector field.

For example in Schwarzschild spacetime the vector field ∂_t is a Killing vector field and it is timelike in the outer part $\mathbb{R} \times (2m, \infty) \times S^2$. In Kerr spacetime the vector field ∂_t is also a Killing vector field. On which subset of $\mathrm{Kerr}_{m,a}$ is ∂_t timelike?
We have to check where $g_{tt} < 0$ holds. Using that $\cos^2\vartheta \leq 1$ we estimate

$$g_{tt} = \frac{-r^2 - a^2\cos^2\vartheta + 2mr}{\rho^2} \geq \frac{-r^2 - a^2 + 2mr}{\rho^2} = -\frac{\Delta}{\rho^2}.$$

On the Boyer-Lindquist block II we have $\Delta < 0$ and thus $g_{tt} > 0$. We expect that at some points of the Boyer-Lindquist blocks I and III we also have $g_{tt} > 0$.

Definition 4.14. The subsets

$$\mathcal{E} := \{p \in I \,|\, g_{tt}(p) > 0\}, \quad \mathcal{E}' := \{p \in III \,|\, g_{tt}(p) > 0\}$$

are called *ergospheres*.

On the intersection $(I \cup III) \cap A$ with the axis A we have $\cos\vartheta = \pm 1$ and therefore $g_{tt} = -\frac{\Delta}{\rho^2} < 0$. Thus we have $\mathcal{E} \cap A = \emptyset$ and $\mathcal{E}' \cap A = \emptyset$. On the equatorial hyperplane Eq we have $\cos\vartheta = 0$ and therefore $g_{tt} = \frac{r(2m-r)}{\rho^2}$ and this is positive if and only if $r \in (0, 2m)$. In a surface of constant t and φ the ergospheres can be pictured as follows (the radius in the picture is an exponential function of r).

Remark 4.15. The submanifolds $I \setminus \overline{\mathcal{E}}$ and $III \setminus \overline{\mathcal{E}'}$ are stationary since ∂_t is a timelike Killing vector field on these submanifolds. We define

$$L := \{p \in \mathrm{Kerr}_{m,a} \,|\, g_{tt}(p) = 0\}.$$

Lemma 4.16. *We have $L = \partial\mathcal{E} \cup \partial\mathcal{E}'$ and L is a smooth timelike hypersurface of* $\mathrm{Kerr}_{m,a}$.

Proof. (a) We write $g_{tt} = \rho^{-2} f$ with

$$f(r, \vartheta) = -r^2 - a^2\cos^2\vartheta + 2mr.$$

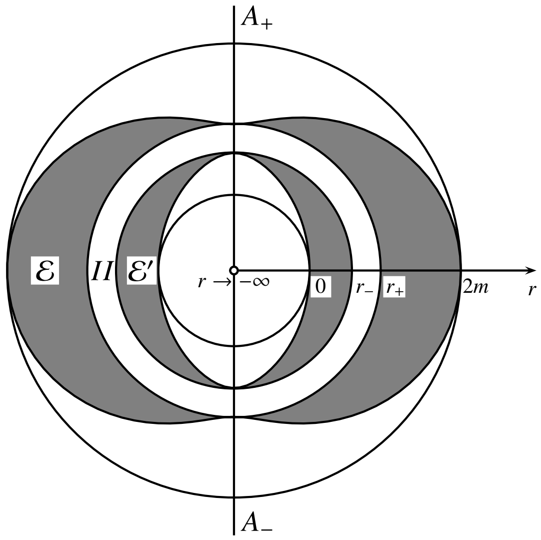


Figure 84.. Ergospheres

Then we have $L = f^{-1}(0)$ and

$$df = 2(m - r)dr + 2a^2 \cos\vartheta \sin\vartheta d\vartheta.$$

If we have $df|_p = 0$ for some $p \in \mathrm{Kerr}_{m,a}$, then $r = m$ and thus $p \in II$. Thus df is nowhere zero on L. This implies that L is a smooth hypersurface of $\mathrm{Kerr}_{m,a}$ and that f changes its sign at L. In particular, in every neighborhood of L one can find points p, q with $g_{tt}(p) > 0$ and $g_{tt}(q) < 0$. It follows that $L \subset \partial\mathcal{E} \cup \partial\mathcal{E}'$. The inclusion $L \supset \partial\mathcal{E} \cup \partial\mathcal{E}'$ is clear by definition of $\mathcal{E}$ and $\mathcal{E}'$.

(b) It remains to show that the induced metric on L has Lorentzian signature. First we note that $L \cap A = \emptyset$ since on A we have $\cos^2\vartheta = 1$ and hence $f = (2m - r)r - a^2 = -\Delta \neq 0$. Thus the vector field ∂_ϑ can be defined on all of L. Clearly the vector fields $\partial_t, \partial_r, \partial_\varphi$ can also be defined on all of L. Since we have

$$df(\partial_t) = df(\partial_\varphi) = 0$$

we conclude that ∂_t and ∂_φ are tangential to L. By Lemma 4.9 we have on L

$$g_{tt}g_{\varphi\varphi} - g_{t\varphi}^2 = -\Delta \sin^2\vartheta < 0$$

and thus ∂_t and ∂_φ generate a 2-dimensional subspace of T_pL with Lorentzian signature. It remains to find a spacelike vector field on L which is perpendicular to this subspace. We set

$$V := a^2 \cos\vartheta \sin\vartheta \partial_r + (r - m)\partial_\vartheta.$$

We easily compute that $df(V) = 0$ and thus we have $V \in T_pL$. Furthermore V is everywhere orthogonal to ∂_t and ∂_φ. Using the coefficients of the metric $g_{(m,a)}$ we compute

$$\langle V, V\rangle = a^4 \cos^2\vartheta \sin^2\vartheta \frac{\rho^2}{\Delta} + (r - m)^2\rho^2 > 0.$$

Thus V is spacelike and T_pL has Lorentzian signature. $\square$

4.2.4. Causality properties and the time machine

We now define a time orientation on the Boyer-Lindquist block I. First we equip $I \setminus \overline{\mathcal{E}}$ with the time orientation for which ∂_t is future-directed. From the coefficients of the metric $g_{(m,a)}$ we see immediately that for every $t_0 \in \mathbb{R}$ the hypersurface $\{t_0\} \times (r_+, \infty) \times S^2$ of I is spacelike. It follows that the vector field grad$t$ is timelike on all of I (note that this does not contradict the fact that ∂_t is not everywhere timelike on I). Furthermore on $I \setminus \overline{\mathcal{E}}$ we have

$$\langle \mathrm{grad} t, \partial_t \rangle = dt(\partial_t) = 1 > 0$$

and thus grad$t$ is past-directed. We extend the time orientation to all of I by requiring that grad$t$ should be past-directed everywhere. Therefore if $\alpha\colon J \to I$ is a causal curve defined on some interval $J \subset \mathbb{R}$, then α is future-directed if and only if

$$0 < \langle \alpha', \mathrm{grad} t \rangle = (t \circ \alpha)' \tag{9}$$

on J. This characterization holds even on $\mathcal{E}$ where ∂_t is spacelike.
Now we want to investigate what happens to a massive particle or to a photon in I as it enters the ergosphere $\mathcal{E}$ and approaches the horizon H_+.

Proposition 4.17. *(1) Let $\alpha\colon J \to \mathcal{E}$ or $\alpha\colon J \to \mathcal{E}'$ be a causal curve. Then for all $s \in J$ we have*

$$(\varphi \circ \alpha)'(s) \neq 0.$$

(2) Let $\alpha\colon J \to \mathcal{E}$ be a future-directed causal curve. Then for all $s \in J$ we have

$$(\varphi \circ \alpha)'(s) > 0.$$

(3) Let $\alpha\colon (s_0, s_1) \to \mathcal{E}$ be a future-directed causal curve with $(r \circ \alpha)(s) \to r_+$, $(\vartheta \circ \alpha)(s) \to \vartheta_0 \in (0, \pi)$ and $(r \circ \alpha)'(s) \to r^* \neq 0$ as $s \nearrow s_1$. Then, as $s \nearrow s_1$, we have

$$(\varphi \circ \alpha)'(s) \to \infty \text{ and } (t \circ \alpha)'(s) \to \infty.$$

Here we have identified the map $\varphi\colon S^2 \setminus \{(0,0,\pm 1)\}$ with a lift $\varphi\colon S^2 \setminus \{(0,0,\pm 1)\} \to \mathbb{R}$. The condition $(\varphi \circ \alpha)' > 0$ then denotes a positively oriented tangent vector to S^1. By the first two assertions in this proposition a particle or a photon entering the ergospheres is forced to rotate in a direction determined by the parameter a. This suggests that a should indeed be interpreted physically as angular momentum per unit mass.

Proof. Ad **(1)**. For fixed $\varphi_0 \in S^1$ we consider the hypersurface $N_{\varphi_0} := \{(t, r, \sigma) \in \mathrm{Kerr}_{m,a} \setminus A \mid \varphi(\sigma) = \varphi_0\}$. At every point of $p \in N_{\varphi_0}$ the vectors ∂_t, ∂_r and ∂_ϑ form a basis of $T_p N_{\varphi_0}$. We recall that ∂_ϑ is spacelike on $\mathrm{Kerr}_{m,a} \setminus A$, ∂_r is spacelike on $I \cup III$ and ∂_t is spacelike on $\mathcal{E} \cup \mathcal{E}'$. It follows that along the trace of α, the hypersurface N_{φ_0} is spacelike. Thus $\alpha'(s) \notin T_{\alpha(s)} N_{\varphi_0}$ for all $s \in J$. This implies that $(\varphi \circ \alpha)'(s) \neq 0$ for all s.

Ad **(2)**. Since α is causal we have $\langle \alpha'(s), \alpha'(s)\rangle \le 0$ for every $s \in J$ and thus

$$0 \ge g_{tt}(\alpha)(t\circ\alpha)'^2 + g_{rr}(\alpha)(r\circ\alpha)'^2 + g_{\vartheta\vartheta}(\alpha)(\vartheta\circ\alpha)'^2 + g_{\varphi\varphi}(\alpha)(\varphi\circ\alpha)'^2 \\ + 2g_{t\varphi}(\alpha)(t\circ\alpha)'(\varphi\circ\alpha)'. \tag{10}$$

Recall that on $\mathcal{E}$ we have $g_{tt} > 0$, $g_{rr} > 0$, $g_{\vartheta\vartheta} > 0$, $g_{\varphi\varphi} > 0$ and $(\varphi\circ\alpha)'^2 > 0$ by **(1)**. Thus we obtain

$$0 > 2g_{t\varphi}(\alpha)(t\circ\alpha)'(\varphi\circ\alpha)'.$$

Since on $\mathcal{E}$ we have $g_{t\varphi} < 0$ and $(t\circ\alpha)' > 0$ by (9), we conclude that $(\varphi\circ\alpha)' > 0$.
Ad **(3)**. Since on $\mathcal{E}$ we have $g_{tt} > 0$, $g_{\vartheta\vartheta} > 0$, $g_{\varphi\varphi} > 0$, $g_{t\varphi} < 0$ and using (10) we obtain

$$\begin{aligned} g_{rr}(\alpha)(r\circ\alpha)'^2 &\le g_{rr}(\alpha)(r\circ\alpha)'^2 + g_{\vartheta\vartheta}(\alpha)(\vartheta\circ\alpha)'^2 + g_{\varphi\varphi}(\alpha)(\varphi\circ\alpha)'^2 \\ &\le -2g_{t\varphi}(\alpha)(t\circ\alpha)'(\varphi\circ\alpha)' - g_{tt}(\alpha)(t\circ\alpha)'^2 \\ &= 2\frac{|g_{t\varphi}(\alpha)|}{\sqrt{g_{tt}(\alpha)}}(\varphi\circ\alpha)'\sqrt{g_{tt}(\alpha)}(t\circ\alpha)' - g_{tt}(\alpha)(t\circ\alpha)'^2 \\ &= -\Big(\frac{|g_{t\varphi}(\alpha)|}{\sqrt{g_{tt}(\alpha)}}(\varphi\circ\alpha)' - \sqrt{g_{tt}(\alpha)}(t\circ\alpha)'\Big)^2 + \frac{g_{t\varphi}(\alpha)^2}{g_{tt}(\alpha)}(\varphi\circ\alpha)'^2 \\ &\le \frac{g_{t\varphi}(\alpha)^2}{g_{tt}(\alpha)}(\varphi\circ\alpha)'^2. \end{aligned} \tag{11}$$

Using $r^2 \le \rho^2$ we estimate

$$\begin{aligned} \frac{g_{t\varphi}^2}{g_{tt}} &= \frac{4m^2r^2a^2\sin^4\vartheta}{\rho^4}\,\frac{\rho^2}{-\rho^2+2mr} = \frac{4m^2r^2a^2\sin^4\vartheta}{\rho^2(-r^2+2mr-a^2\cos^2\vartheta)} = \frac{4m^2r^2a^2\sin^4\vartheta}{\rho^2(a^2\sin^2\vartheta-\Delta)} \\ &\le \frac{4m^2a^2\sin^4\vartheta}{a^2\sin^2\vartheta-\Delta}. \end{aligned}$$

We evaluate this expression at $\alpha(s)$ and let $s \nearrow s_1$. Then $\Delta(\alpha(s)) \to 0$ because of $r(\alpha(s)) \to r_+$ and thus the right hand side tends to $4m^2\sin^2\vartheta_0$. Thus for s close to s_1 we have $\frac{g_{t\varphi}(\alpha)^2}{g_{tt}(\alpha)} \le 8m^2\sin^2\vartheta_0$ and therefore

$$g_{rr}(\alpha)(r\circ\alpha)'^2 \le 8m^2\sin^2\vartheta_0(\varphi\circ\alpha)'^2 \le 8m^2(\varphi\circ\alpha)'^2.$$

Furthermore as $s \nearrow s_1$ we have by hypothesis

$$(r\circ\alpha)'(s)^2 \to (r^*)^2 > 0 \text{ and } g_{rr}(\alpha(s)) \to \infty.$$

Thus as $s \nearrow s_1$ we obtain $(\varphi\circ\alpha)'(s)^2 \to \infty$ and by **(2)** we find $(\varphi\circ\alpha)'(s) \to \infty$.
Next we can repeat the estimate (11) with the roles of the coordinates t and φ interchanged and we get

$$g_{rr}(\alpha)(r\circ\alpha)'^2 \le \frac{g_{t\varphi}(\alpha)^2}{g_{\varphi\varphi}(\alpha)}(t\circ\alpha)'^2.$$

Since in I we have $2mra^2 \sin^2 \vartheta \geq 0$ we estimate

$$\frac{g_{t\varphi}^2}{g_{\varphi\varphi}} = \frac{4m^2r^2a^2 \sin^4 \vartheta}{\rho^4} \frac{\rho^2}{(\rho^2(r^2 + a^2) + 2mra^2 \sin^2 \vartheta) \sin^2 \vartheta} \leq \frac{4m^2r^2a^2 \sin^2 \vartheta}{\rho^4(r^2 + a^2)} \leq \frac{4m^2a^2}{\rho^4}.$$

If we evaluate this expression at $\alpha(s)$ and let $s \nearrow s_1$ the right hand side remains bounded. Thus there exists $C_1 > 0$ such that for all s close to s_1 we have

$$g_{rr}(\alpha)(r \circ \alpha)'^2 \leq C_1(t \circ \alpha)'^2.$$

As above we get $(t\circ\alpha)'(s)^2 \to \infty$ as $s \nearrow s_1$. Since α is future-directed we get $(t\circ\alpha)'(s) \to \infty$. □

If there exists a closed timelike curve in a spacetime then this model of the universe predicts an observer who can influence his own past. This is then interpreted as a violation of causality. We want to investigate whether this can happen in Kerr spacetime.

Proposition 4.18. *In the Boyer-Lindquist blocks I and II there exist no closed causal curves.*

Proof. (a) Let $\alpha\colon J \to I$ be a causal curve. Without loss of generality we may assume that α is future-directed. By (9) we have $(t \circ \alpha)' > 0$. Thus α is injective and in particular not closed.

(b) Let $\alpha\colon J \to II$ be a causal curve. For fixed $r_0 \in (r_-, r_+)$ we consider the hypersurface $N_{r_0} := \mathbb{R} \times \{r_0\} \times S^2 \subset II$. At every point $p \in N_{r_0} \setminus A$ the vector fields $\partial_t, \partial_\vartheta, \partial_\varphi$ form a basis of $T_pN_{r_0}$. Since $\langle\partial_r, \partial_t\rangle = \langle\partial_r, \partial_\vartheta\rangle = \langle\partial_r, \partial_\varphi\rangle = 0$ the coordinate field ∂_r is perpendicular to $N_{r_0} \setminus A$. By continuity, this holds along all of N_{r_0}.
The vector field ∂_r is timelike on II and thus the hypersurface N_{r_0} is spacelike. Since α is a causal curve we have $\alpha'(s) \notin T_{\alpha(s)}N_{r_0}$ for all $s \in J$ and all $r_0 \in (r_-, r_+)$. It follows that $(r \circ \alpha)'$ is nowhere zero. Thus α is injective and in particular not closed. □

In order to prepare the discussion of the Boyer-Lindquist block III we introduce the following vector fields.

Definition 4.19. The *canonical vector fields* on $\mathrm{Kerr}_{m,a}$ are defined by

$$V := (r^2 + a^2)\partial_t + a\partial_\varphi, \quad W := \partial_\varphi + a \sin^2 \vartheta \partial_t.$$

Note that $V$and W are well-defined smooth vector fields on all of $\mathrm{Kerr}_{m,a}$ since $\sin^2 \vartheta$ and ∂_φ are defined and smooth on all of S^2.

Lemma 4.20. *We have*

$$g_{(m,a)}(V,V) = -\Delta\rho^2, \quad g_{(m,a)}(W,W) = \rho^2 \sin^2\vartheta, \quad g_{(m,a)}(V,W) = 0.$$

Proof. This follows immediately from the Boyer-Lindquist identities (BL1)-(BL4). □

Remark 4.21. At every point of $\mathrm{Kerr}_{m,a} \setminus A$ the vector fields V, W, ∂_r, ∂_ϑ form an orthogonal basis of the tangent space. The vector field V is timelike on $I \cup III$ and spacelike on II. On I we have $\langle \partial_t, V\rangle = (r^2 + a^2)g_{tt} + ag_{t\varphi} = -\Delta < 0$, i.e. V is future-directed there. We choose the time orientation on the Boyer-Lindquist block III for which V is future directed there too.

Definition 4.22. The region $C := \{p \in \mathrm{Kerr}_{m,a} \,|\, g_{\varphi\varphi}(p) < 0\} \subset III$ is called the *Carter time machine*.

Remark 4.23. On the equatorial hyperplane Eq we have $\sin^2\vartheta = 1$ and thus

$$g_{\varphi\varphi} < 0 \iff r^2 + a^2 + \frac{2mra^2}{r^2} < 0 \iff r^4 + a^2r^2 + 2mra^2 < 0.$$

The polynomial $f(r) = r^4 + a^2r^2 + 2ma^2r$ has a simple root at $r = 0$ and thus changes its sign at $r = 0$. Therefore $f(r) < 0$ for negative r close to 0. In particular, the subset $C \subset III$ is not empty.

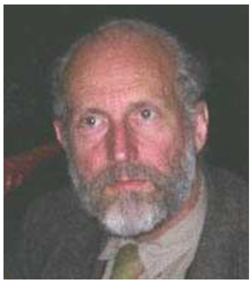


Figure 85.. Brandon Carter (*1942)

In a surface of constant t and φ, the Carter time machine can be pictured as follows (the radius in the picture is an exponential function of r). We claim that for any two points $p, q \in III$ we

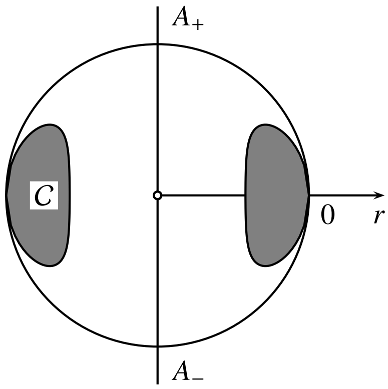


Figure 86.. Carter time machine

can find a future-directed timelike curve from p to q in III using the Carter time machine. In the following we will construct such a curve.

Step 1: We find a curve from p to some point in C. To this end, we fix $\bar{r} < 0$ such that $\mathbb{R} \times \{\bar{r}\} \times \{\sigma \in S^2 \,|\, \vartheta(\sigma) = \frac{\pi}{2}\}$ is contained in C. We write $p = (t_0, r_0, \sigma_0) \in \mathbb{R}^2 \times S^2 \setminus \Sigma$ and consider the functions

$$\begin{aligned} r_1 : \quad & [0,1] \to (-\infty, r_-), \quad r_1(s) := s\bar{r} + (1-s)r_0, \\ \sigma_1 : \quad & [0,1] \to S^2, \quad \sigma_1(0) = \sigma_0,\ \vartheta(\sigma_1(1)) = \tfrac{\pi}{2}, \end{aligned}$$

where σ_1 is a shortest geodesic on S^2 from σ_0 to the equator. It satisfies $\varphi_0 := \varphi \circ \sigma_1 = \text{const}$ and $(\vartheta \circ \sigma_1)' = \text{const}$. Furthermore, let $\tau_1 \colon [0,1] \to \mathbb{R}$ be the function satisfying $\tau_1'(s) = r_1(s)^2 + a^2$ and $\tau_1(0) = 0$. For a constant $c_1 \in \mathbb{R}$ we consider the curve $\gamma_1 \colon [0,1] \to \mathrm{Kerr}_{m,a}$,

$$\gamma_1(s) := (t_0 + c_1\tau_1(s), r_1(s), R_{c_1 as}\sigma_1(s))$$

where $R_{c_1 as}$ is the rotation by the angle of $c_1 as$ around the z-axis. In particular, we have $\varphi(R_{c_1 as}\sigma_1(s)) = \varphi_0 + c_1 as$. It follows that

$$\begin{aligned} \gamma_1'(s) &= c_1(r_1(s)^2 + a^2)\partial_t + r_1'(s)\partial_r + (\vartheta \circ \sigma_1)'(s)\partial_\vartheta + c_1 a\partial_\varphi \\ &= c_1 V + r_1'(s)\partial_r + (\vartheta \circ \sigma_1)'(s)\partial_\vartheta \end{aligned}$$

where V is one of the canonical vector fields. By Lemma 4.20 we have

$$\langle \gamma_1'(s), \gamma_1'(s)\rangle = -c_1^2 \underbrace{\Delta\rho^2}_{\geq c_3} + \underbrace{r_1'(s)^2\frac{\rho^2}{\Delta} + (\vartheta \circ \sigma_1)'(s)^2\rho^2}_{\leq c_2}$$

where $c_2, c_3 > 0$ depend only on p and $\bar{r}$. Therefore if $c_1 > (\frac{c_2}{c_3})^{1/2}$ the curve γ_1 is timelike. Using Lemma 4.20 we obtain

$$\langle \gamma_1'(s), V\rangle = c_1\langle V, V\rangle = -c_1\Delta\rho^2 < 0$$

and thus by our definition of time-orientation on III the curve γ_1 is future-directed. We denote its endpoint by

$$\gamma_1(1) =: (t_0 + c_1\tau_1(1), \bar{r}, u_1) =: (t_1, \bar{r}, u_1), \quad \vartheta(u_1) = \tfrac{\pi}{2}.$$

Remark 4.24. Had we chosen $c_1 < -(\frac{c_2}{c_3})^{1/2}$, we would have obtained a past-directed timelike curve from p to some point in $\mathbb{R} \times \{\bar{r}\} \times \{\sigma \in S^2 \,|\, \vartheta(\sigma) = \frac{\pi}{2}\}$.

Step 2: Let $t_1, t_2 \in \mathbb{R}$, $u_1, u_2 \in S^2$, $\vartheta(u_1) = \vartheta(u_2) = \frac{\pi}{2}$. Given a constant $c_4 > 0$ we define the curve $\gamma_2 \colon [0,1] \to III$,

$$\gamma_2(s) := (t_1 + s(t_2 - t_1), \bar{r}, R_{-c_4 s}u_1).$$

We compute

$$\gamma_2'(s) = (t_2 - t_1)\partial_t - c_4\partial_\varphi.$$

Thus if c_4 is large enough, γ_2' is timelike because ∂_φ is timelike on C. Using the Boyer-Lindquist identities (BL3) and (BL4) we obtain

$$\begin{aligned}\langle\gamma_2'(s), V\rangle &= \langle(t_2-t_1)\partial_t - c_4\partial_\varphi, (\overline{r}^2+a^2)\partial_t + a\partial_\varphi\rangle\\ &= (t_2-t_1)(\overline{r}^2+a^2)g_{tt} - c_4 a g_{\varphi\varphi} + ((t_2-t_1)a - c_4(\overline{r}^2+a^2))g_{t\varphi}\\ &= -(t_2-t_1)\Delta - c_4\Delta a\sin^2\vartheta\\ &= -\Delta(c_4 a + t_2 - t_1).\end{aligned}$$

If c_4 is large enough this is negative and thus γ_2 is a future-directed timelike curve from $(t_1, \overline{r}, u_1)$ to $(t_2, \overline{r}, \tilde{u})$ where $\tilde{u}$ is some point on the equator in S^2. By enlarging c_4 further we can arrange that $\tilde{u}$ coincides with the given point u_2.
Step 3: By Remark 4.24 there exists a past-directed timelike curve $\hat{\gamma}_3\colon [0,1] \to III$ with $\hat{\gamma}_3(0) = q$ and $\hat{\gamma}_3(1) = (t_2, \overline{r}, u_2)$ for some $t_2 \in \mathbb{R}$, $u_2 \in S^2$, $\vartheta(u_2) = \frac{\pi}{2}$. We define $\gamma_3(s) := \hat{\gamma}_3(1-s)$, $s \in [0,1]$.
For any points $p, q \in III$ we have therefore found a future-directed timelike curve from p to q. In particular, we see that unlike in the Boyer-Lindquist blocks I and II there exist many closed causal curves in III.

4.2.5. Extension across the horizons

Next we claim that we can extend the metric $g_{(m,a)}$ across the horizon H. To this end, we consider the diffeomorphism

$$\mathrm{Kerr}_{m,a} \to \mathrm{Kerr}_{m,a}, \quad (t, r, \sigma) \mapsto (t + \xi(r), r, R_{\eta(r)}\sigma) =: (t^*, r^*, \sigma^*)$$

where $\xi, \eta\colon \mathbb{R}\setminus\{r_-, r_+\} \to \mathbb{R}$ are functions such that

$$\frac{d\xi}{dr} = \frac{r^2+a^2}{\Delta(r)}, \qquad \frac{d\eta}{dr} = \frac{a}{\Delta(r)}.$$

and $R_{\eta(r)}$ is the rotation by the angle $\eta(r)$ around the z-axis. Note that on each of the intervals $(-\infty, r_-)$, (r_-, r_+), (r_+, ∞) the functions ξ and η are only defined up to an additive constant. It is clear that the above map is a diffeomorphism since the inverse map is

$$(t^*, r^*\sigma^*) \mapsto (t^* - \xi(r^*), r^*, R_{-\eta(r^*)}\sigma^*)\,.$$

In other words, we consider the transformation of coordinates

$$t^* = t + \xi(r), \quad r^* = r, \quad \vartheta^* = \vartheta, \quad \varphi^* = \varphi + \eta(r).$$

We compute the Kerr metric pulled back along this diffeomorphism, i.e. in the new coordinates t^*, r^*, ϑ^*, φ^*. This coordinate system is known as *Kerr coordinates*.

Lemma 4.25. *In Kerr coordinates $(t^*, r^*, \vartheta^*, \varphi^*)$ on* $\mathrm{Kerr}_{m,a} \setminus A$ *we have*

$$g_{(m,a)} = g_{tt}(dt^*)^2 + \rho^2(d\vartheta^*)^2 + g_{\varphi\varphi}(d\varphi^*)^2 + 2dt^*dr^* - 2a\sin^2\vartheta dr^*d\varphi^* + 2g_{t\varphi}dt^*d\varphi^*.$$

Proof. By definition we have

$$dt^* = dt + \xi'(r)dr = dt + \frac{r^2+a^2}{\Delta}dr, \quad dr^* = dr, \quad d\vartheta^* = d\vartheta, \quad d\varphi^* = d\varphi + \frac{a}{\Delta}dr.$$

It follows that

$$\begin{aligned} g_{(m,a)} =& g_{tt}\Big(dt^* - \frac{r^2+a^2}{\Delta}dr^*\Big)^2 + g_{rr}(dr^*)^2 + g_{\vartheta\vartheta}(d\vartheta^*)^2 + g_{\varphi\varphi}\Big(d\varphi^* - \frac{a}{\Delta}dr^*\Big)^2 \\ &+ 2g_{t\varphi}\Big(dt^* - \frac{r^2+a^2}{\Delta}dr^*\Big)\Big(d\varphi^* - \frac{a}{\Delta}dr^*\Big) \\ =& g_{tt}(dt^*)^2 + \Big(g_{tt}\Big(\frac{r^2+a^2}{\Delta}\Big)^2 + \frac{\rho^2}{\Delta} + g_{\varphi\varphi}\frac{a^2}{\Delta^2} + 2g_{t\varphi}\frac{(r^2+a^2)a}{\Delta^2}\Big)(dr^*)^2 + \rho^2(d\vartheta^*)^2 \\ &+ g_{\varphi\varphi}(d\varphi^*)^2 - 2\Big(g_{tt}\frac{r^2+a^2}{\Delta} + g_{t\varphi}\frac{a}{\Delta}\Big)dt^*dr^* - 2\Big(g_{\varphi\varphi}\frac{a}{\Delta} + g_{t\varphi}\frac{r^2+a^2}{\Delta}\Big)dr^*d\varphi^* \\ &+ 2g_{t\varphi}dt^*d\varphi^*. \end{aligned}$$

By the Boyer-Lindquist identities (BL3) and (BL4) the coefficient of $(dr^*)^2$ is equal to

$$\frac{r^2+a^2}{\Delta^2}(-\Delta) + \frac{a}{\Delta^2}\Delta a\sin^2\vartheta + \frac{\rho^2}{\Delta} = \frac{-r^2-a^2+a^2\sin^2\vartheta}{\Delta} + \frac{\rho^2}{\Delta} = 0.$$

By (BL4) the coefficient of dt^*dr^* is equal to 2 and by (BL3) the coefficient of $dr^*d\varphi^*$ is equal to $-2a\sin^2\vartheta$. This completes the proof. □

The term dr^2 has disappeared in these new coordinates. All coefficient functions in this formula for the metric are smooth even on H since the denominator Δ is no longer present. Thus we obtain a smooth extension of $g_{(m,a)}$ as a $(0,2)$-tensor field on $\mathrm{Kerr}_{m,a} \cup (H \setminus A)$.
In order to show that this extension has Lorentzian signature, we express the volume form in the new coordinates $t^*, r^*, \vartheta^*, \varphi^*$. By Lemma 4.9 we have

$$\begin{aligned} \mathrm{vol} &= \sqrt{|\det g|}dt \wedge dr \wedge d\vartheta \wedge d\varphi \\ &= \rho^2|\sin\vartheta|(dt^* - \xi dr^*) \wedge dr \wedge d\vartheta \wedge (d\varphi^* - \eta dr) \\ &= \rho^2|\sin\vartheta|dt^* \wedge dr^* \wedge d\vartheta^* \wedge d\varphi^* \end{aligned}$$

and thus $\mathrm{vol} \neq 0$ on $H \setminus A$. It follows that the extension of $g_{(m,a)}$ is non-degenerate on $H \setminus A$ and therefore must be Lorentzian on $H \setminus A$ by continuity.

In order to show that we also obtain an extension to $H \cap A$, we first note that on $S^2 \setminus \{(0,0,\pm 1)\}$ we have $g_{S^2} = d\vartheta^2 + \sin^2\vartheta d\varphi^2$. Thus on $H \setminus A$ we obtain

$$g_{(m,a)} = g_{tt}(dt^*)^2 + \rho^2 g_{S^2} + (-\rho^2\sin^2\vartheta + g_{\varphi\varphi})(d\varphi^*)^2 + 2dt^*dr^* - 2a(\sin^2\vartheta d\varphi^*)dr^*$$
$$- \frac{4mra}{\rho^2}(\sin^2\vartheta d\varphi^*)dt^*.$$

For the third term on the right hand side we obtain

$$\Big((r^2 + a^2 - \rho^2)\sin^2\vartheta + \frac{2mra^2}{\rho^2}\sin^4\vartheta\Big)(d\varphi^*)^2 = \Big(a^2 + \frac{2mra^2}{\rho^2}\Big)(\sin^2\vartheta d\varphi^*)^2.$$

By (7) the 1-form $\sin^2\vartheta d\varphi^*$ extends smoothly to A and thus we have obtained a smooth extension of $g_{(m,a)}$ to $H \cap A$ as a $(0,2)$-tensor field. On $H \cap A$ we have

$$g_{(m,a)} = g_{tt}(dt^*)^2 + \rho^2 g_{S^2} + 2dt^*dr^*.$$

Thus since

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

and $\rho^2 g_{S^2}$ is Riemannian the extension of $g_{(m,a)}$ is Lorentzian on $H \cap A$. Therefore we have extended the Lorentzian metric $g_{(m,a)}$ to

$$\mathrm{Kerr}^*_{m,a} := \mathrm{Kerr}_{m,a} \cup H = (\mathbb{R}^2 \times S^2) \setminus \Sigma.$$

Next we examine the geometry of the horizon H.

(1) The manifolds $H_\pm = \mathbb{R} \times \{r_\pm\} \times S^2$ are diffeomorphic to $\mathbb{R} \times S^2$.

(2) The induced metric on $H_\pm \setminus A$ in the coordinates $t^*, \vartheta^*, \varphi^*$ has determinant

$$\det\begin{pmatrix} g_{tt} & 0 & g_{t\varphi} \\ 0 & \rho^2 & 0 \\ g_{t\varphi} & 0 & g_{\varphi\varphi} \end{pmatrix} = \rho^2 \det\begin{pmatrix} g_{tt} & g_{t\varphi} \\ g_{t\varphi} & g_{\varphi\varphi} \end{pmatrix} = -\Delta\rho^2\sin^2\vartheta = 0$$

where we have used Lemma 4.9. Since the metric is continuous this holds on all of $H_\pm$ and thus $H_\pm$ are lightlike hypersurfaces of $\mathrm{Kerr}^*_{m,a}$.

(3) We claim that $H_\pm$ are totally geodesic hypersurfaces of $\mathrm{Kerr}^*_{m,a}$. Before we come to the proof we first recall that the canonical vector field $V = (r^2 + a^2)\partial_t + a\partial_\varphi$ is defined on $\mathrm{Kerr}_{m,a}$. We express V in the coordinates $t^*, r^*, \vartheta^*, \varphi^*$:

$$dt^*(V) = \Big(dt + \frac{r^2 + a^2}{\Delta}dr\Big)((r^2 + a^2)\partial_t + a\partial_\varphi) = r^2 + a^2,$$
$$d\varphi^*(V) = (d\varphi + \frac{a}{\Delta}dr)((r^2 + a^2)\partial_t + a\partial_\varphi) = a,$$
$$dr^*(V) = dr(V) = 0,$$
$$d\vartheta^*(V) = \vartheta(V) = 0,$$

hence

$$V = (r^2 + a^2)\partial_{t^*} + a\partial_{\varphi^*} .$$

Thus V extends to a smooth vector field on $\mathrm{Kerr}^*_{m,a}$. By continuity, the formula

$$\langle V, V\rangle = -\Delta\rho^2$$

from Lemma 4.20 holds on all of $\mathrm{Kerr}^*_{m,a}$. Therefore V is lightlike along $H_\pm$. Since it is tangential to $H_\pm$ this implies that it is also normal to $H_\pm$. In other words, we have for $p \in H_\pm$

$$T_pH_\pm = \{X \in T_p\mathrm{Kerr}^*_{m,a} \,|\, \langle X, V\rangle = 0\}. \tag{12}$$

Next we define the vector fields

$$V_\pm := (r_\pm^2 + a^2)\partial_{t*} + a\partial_{\varphi^*} = (r_\pm^2 + a^2)\partial_t + a\partial_\varphi$$

on $\mathrm{Kerr}^*_{m,a}$. Since ∂_t and ∂_φ are Killing vector fields on $\mathrm{Kerr}_{m,a}$ so are $V_\pm$. By continuity, $V_\pm$ are Killing vector fields on all of $\mathrm{Kerr}^*_{m,a}$. Moreover, we have $V_\pm = V$ along $H_\pm$.
Now we can prove that $H_\pm$ are totally geodesic hypersurfaces of $\mathrm{Kerr}^*_{m,a}$. To this end, let X, Y be vector fields on $\mathrm{Kerr}^*_{m,a}$ which are tangential to $H_\pm$ along $H_\pm$. Then along $H_\pm$ we have

$$\langle\nabla_X Y, V\rangle = \langle\nabla_X Y, V_\pm\rangle = \partial_X \underbrace{\langle Y, V_\pm\rangle}_{\equiv 0} - \langle Y, \nabla_X V_\pm\rangle = \langle X, \nabla_Y V_\pm\rangle$$

where in the last step we have used that $V_\pm$ are Killing vector fields. By exchanging X and Y we get

$$\langle\nabla_Y X, V_\pm\rangle = -\langle X, \nabla_Y V_\pm\rangle.$$

Since X, Y are tangential to $H_\pm$ the bracket $[X, Y]$ is tangential to $H_\pm$ along $H_\pm$ and thus

$$0 = \langle[X, Y], V_\pm\rangle = \langle\nabla_X Y - \nabla_Y X, V_\pm\rangle = 2\langle X, \nabla_Y V_\pm\rangle = 2\langle\nabla_X Y, V_\pm\rangle.$$

By (12) it follows that $\nabla_X Y$ is tangential to $H_\pm$. Thus $H_\pm$ are totally geodesic hypersurfaces of $\mathrm{Kerr}^*_{m,a}$.
(4) We claim that the integral curves of V along $H_\pm$ are pregeodesics. Namely let $\gamma\colon J \to \mathrm{Kerr}^*_{m,a}$ be a geodesic with $\gamma(0) \in H_\pm$ and $\gamma'(0) = V(\gamma(0))$. Then $\gamma'(0)$ is lightlike and since γ is a geodesic we know that $\gamma'(s)$ is lightlike for all $s \in J$. Since $H_\pm$ are totally geodesic hypersurfaces the geodesic γ does not leave $H_\pm$, i.e. we have $\gamma'(s) \in T_{\gamma(s)}H_\pm$ for all $s \in J$. By Lemma 1.12, all vectors in any tangent space $T_pH \setminus \mathbb{R}\cdot V(p)$ are spacelike. Thus there exists a smooth function $\alpha\colon J \to \mathbb{R}$ such that $\gamma'(s) = \alpha(s)V(\gamma(s))$ for all $s \in J$. Hence the integral curve of V through $\gamma(0)$ is a reparametrization of the geodesic γ.

Next we define a time orientation on $\mathrm{Kerr}^*_{m,a}$. We have already defined a time orientation on the Boyer-Lindquist blocks I and III such that V is future directed. In the following we want to use the coordinate vector field ∂_{r^*} to define a time orientation on $\mathrm{Kerr}^*_{m,a}$. The vector field ∂_{r^*} is defined on all of $\mathrm{Kerr}^*_{m,a}$. By Lemma 4.25 it is lightlike and we have

$$\langle\partial_{r^*}, V\rangle = (r^2 + a^2)\langle\partial_{r^*}, \partial_{t^*}\rangle + a\langle\partial_{r^*}, \partial_{\varphi^*}\rangle = (r^2 + a^2) - a^2\sin^2\vartheta = \rho^2 > 0.$$

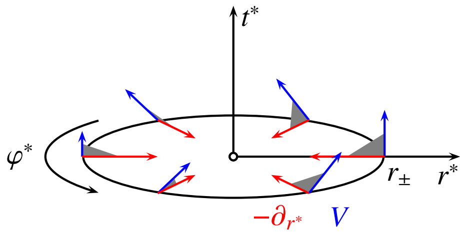


Figure 87.. Future time cones along the horizons

Therefore ∂_{r^*} is past directed on $I \cup III$ and we define a time orientation on $\mathrm{Kerr}^*_{m,a}$ by requiring that ∂_{r^*} should be past directed everywhere. We can imagine the future timecones along $H_\pm$ to be directed towards the inner side of $H_\pm$. Thus if $\alpha\colon J \to \mathrm{Kerr}^*_{m,a}$ is a future-directed timelike curve with $\alpha(s_0) \in H_\pm$, then $(r^* \circ \alpha)'(s_0) = dr^*(\alpha'(s_0)) < 0$. The same holds true for future-directed lightlike curves α provided that $\alpha'(s_0) \notin \mathbb{R} \cdot V(\alpha(s_0))$. This means that massive particles can pass through the horizon from block I to block II or from block II to block III but not in the reverse direction and that photons either stay on the horizon or behave in the same way.

Remark 4.26. The boundaries of the ergospheres $\mathcal{E} \subset I$ and $\mathcal{E}' \subset III$ are submanifolds of $\mathrm{Kerr}^*_{m,a}$. By Lemma 4.16 we know that $\partial\mathcal{E}\backslash A$ and $\partial\mathcal{E}'\backslash A$ are timelike hypersurfaces of $\mathrm{Kerr}^*_{m,a}\backslash A$. For $p \in \partial\mathcal{E} \cap A$ we see that $T_p\partial\mathcal{E} = T_pH$ is lightlike and similarly for $p \in \partial\mathcal{E}' \cap A$. We conclude that massive particles or photons coming from I or III can enter the ergospheres and leave it.

4.2.6. The Christoffel symbols

For the study of geodesics we need to understand the Levi-Civita connection of the Kerr metric, in other words, we have to know the Christoffel symbols. Here they are in Boyer-Lindquist coordinates.

Proposition 4.27. *The Christoffel symbols of $g_{(m,a)}$ in Boyer-Lindquist coordinates are given by*

(a) $\Gamma^r_{tt} = -\frac{\Delta m(\rho^2-2r^2)}{\rho^6}$, $\Gamma^\vartheta_{tt} = -\frac{2mra^2 \sin\vartheta\cos\vartheta}{\rho^6}$, $\Gamma^t_{tt} = \Gamma^\varphi_{tt} = 0$,

(b) $\Gamma^r_{\varphi\varphi} = -\frac{\Delta}{\rho^2}\left(r + \frac{ma^2\sin^2\vartheta(\rho^2-2r^2)}{\rho^4}\right)\sin^2\vartheta$, $\Gamma^t_{\varphi\varphi} = \Gamma^\varphi_{\varphi\varphi} = 0$,
$\Gamma^\vartheta_{\varphi\varphi} = -\frac{\sin\vartheta\cos\vartheta}{\rho^2}\left(r^2 + a^2 + \frac{2mra^2\sin^2\vartheta(2\rho^2+a^2\sin^2\vartheta)}{\rho^4}\right)$,

(c) $\Gamma^r_{t\varphi} = \Gamma^r_{\varphi t} = \frac{\Delta ma\sin^2\vartheta(\rho^2-2r^2)}{\rho^6}$, $\Gamma^\vartheta_{t\varphi} = \Gamma^\vartheta_{\varphi t} = \frac{2mra\sin\vartheta\cos\vartheta(r^2+a^2)}{\rho^6}$, $\Gamma^t_{t\varphi} = \Gamma^t_{\varphi t} = \Gamma^\varphi_{t\varphi} = \Gamma^\varphi_{\varphi t} = 0$,

(d) $\Gamma^t_{rr} = \Gamma^\varphi_{rr} = \Gamma^t_{r\vartheta} = \Gamma^\varphi_{r\vartheta} = \Gamma^t_{\vartheta r} = \Gamma^\varphi_{\vartheta r} = \Gamma^t_{\vartheta\vartheta} = \Gamma^\varphi_{\vartheta\vartheta} = 0$, $\Gamma^r_{rr} = \frac{r}{\rho^2} + \frac{m-r}{\Delta}$, $\Gamma^\vartheta_{rr} = \frac{a^2\sin\vartheta\cos\vartheta}{\Delta\rho^2}$,
$\Gamma^r_{\vartheta\vartheta} = -\frac{\Delta r}{\rho^2}$, $\Gamma^\vartheta_{\vartheta\vartheta} = -\frac{a^2\sin\vartheta\cos\vartheta}{\rho^2}$, $\Gamma^r_{r\vartheta} = \Gamma^r_{\vartheta r} = -\frac{a^2\sin\vartheta\cos\vartheta}{\rho^2}$, $\Gamma^\vartheta_{r\vartheta} = \Gamma^\vartheta_{\vartheta r} = \frac{r}{\rho^2}$,

(e) $\Gamma^r_{rt} = \Gamma^r_{tr} = \Gamma^r_{\vartheta t} = \Gamma^r_{t\vartheta} = 0$,

(f) $\Gamma^r_{r\varphi} = \Gamma^r_{\varphi r} = \Gamma^r_{\vartheta\varphi} = \Gamma^r_{\varphi\vartheta} = 0$, $\Gamma^\varphi_{tr} = \Gamma^\varphi_{rt} = \frac{ma(2r^2-\rho^2)}{\Delta\rho^4}$, $\Gamma^t_{tr} = \Gamma^t_{rt} = \frac{m(r^2+a^2)(2r^2-\rho^2)}{\Delta\rho^4}$,
$\Gamma^t_{\varphi r} = \Gamma^t_{r\varphi} = \frac{ma\sin^2\vartheta}{\Delta\rho^4}((r^2+a^2)(\rho^2-2r^2)-2r^2\rho^2)$, $\Gamma^\varphi_{\varphi r} = \Gamma^\varphi_{r\varphi} = \frac{r\rho^4+ma^2\rho^2\sin^2\vartheta-2mr^2(r^2+a^2)}{\Delta\rho^4}$,

(g) $\Gamma^\vartheta_{rt} = \Gamma^\vartheta_{tr} = \Gamma^\vartheta_{\vartheta t} = \Gamma^\vartheta_{t\vartheta} = 0$,

(h) $\Gamma^\vartheta_{r\varphi} = \Gamma^\vartheta_{\varphi r} = \Gamma^\vartheta_{\vartheta\varphi} = \Gamma^\vartheta_{\varphi\vartheta} = 0$, $\Gamma^\varphi_{t\vartheta} = \Gamma^\varphi_{\vartheta t} = -\frac{2mra\cos\vartheta}{\rho^4\sin\vartheta}$, $\Gamma^t_{t\vartheta} = \Gamma^t_{\vartheta t} = -\frac{2mra^2\sin\vartheta\cos\vartheta}{\rho^4}$,
$\Gamma^\varphi_{\vartheta\varphi} = \Gamma^\varphi_{\varphi\vartheta} = \frac{\cos\vartheta}{\sin\vartheta}\Big(1+\frac{2mra^2\sin^2\vartheta}{\rho^4}\Big)$, $\Gamma^t_{\vartheta\varphi} = \Gamma^t_{\varphi\vartheta} = \frac{2mra^3\sin^3\vartheta\cos\vartheta}{\rho^4}$.

Proof. This is a straighforward but lengthy and tedious computation best left to SageMath, see Section B.2.3. □

Definition 4.28. A future directed timelike curve of the form $\alpha(s) = (t(s), r_0, \sigma_0)$ which is parametrized by proper time is called a *Kerr observer*.

Remark 4.29. For a Kerr observer we have

$$-1 = \langle\alpha', \alpha'\rangle = t'(s)^2 g_{tt} = t'(s)^2\Big(-1+\frac{2mr_0}{\rho_0^2}\Big)$$

with $\rho_0 := \rho(r_0, \sigma_0)$. Thus for sufficiently large r_0 we get $t'(s) = (1-\frac{2mr_0}{\rho_0^2})^{-1/2}$. It follows that

$$\alpha(s) = \Big(s_0 + \Big(1-\frac{2mr_0}{\rho_0^2}\Big)^{-1/2} s, r_0, \sigma_0\Big)$$

and hence

$$\alpha' = \Big(1-\frac{2mr_0}{\rho_0^2}\Big)^{-1/2}\partial_t.$$

Putting $\Delta_0 := \Delta(r_0)$ and using Proposition 4.27 (a) we find for the covariant derivative

$$\begin{aligned}
\frac{\nabla}{ds}\alpha'(s) &= \nabla_{(1-2mr_0/\rho_0^2)^{-1/2}\partial_t}\Big[\Big(1-\frac{2mr_0}{\rho_0^2}\Big)^{-1/2}\partial_t\Big] \\
&= \Big(1-\frac{2mr_0}{\rho_0^2}\Big)^{-1}\nabla_{\partial_t}\partial_t \\
&= \Big(1-\frac{2mr_0}{\rho_0^2}\Big)^{-1}(\Gamma^t_{tt}\partial_t + \Gamma^r_{tt}\partial_r + \Gamma^\vartheta_{tt}\partial_\vartheta + \Gamma^\varphi_{tt}\partial_\varphi) \\
&= \Big(1-\frac{2mr_0}{\rho_0^2}\Big)^{-1}\Big(\frac{\Delta_0 m}{\rho_0^4}\Big(-1+\frac{2r_0^2}{\rho_0^2}\Big)\partial_r - \frac{2mr_0a^2\sin\vartheta\cos\vartheta}{\rho_0^6}\partial_\vartheta\Big).
\end{aligned}$$

As $r_0 \to \infty$ we have $\rho_0^2 \sim r_0^2$ and $\Delta_0 \sim r_0^2$ (in the sense that the quotient tend to 1) and thus

$$\frac{\nabla}{ds}\alpha'(s) \sim \frac{m}{r_0^2}\partial_r - \frac{2ma^2 \sin\vartheta \cos\vartheta}{r_0^5}\partial_\vartheta.$$

For large r_0 a Kerr observer has to use a force approximately proportional to $\frac{m}{r_0^2}$ in order to keep his r-coordinate constant. This coincides with the prediction of Newton's theory of the gravitational force exerted on the observer by a body of mass m. This justifies to call the parameter m the mass of the black hole.

4.2.7. Geodesics

Our next aim is to examine geodesics in $\mathrm{Kerr}^*_{m,a}$. More precisely, we will derive 4 quantities which are conserved along geodesics and which will help to reduce the geodesic equation to a system of 4 differential equations of first order. Let $\gamma\colon J \to \mathrm{Kerr}^*_{m,a}$ be a geodesic. It follows immediately that

$$q := \langle \gamma', \gamma' \rangle$$

is constant on J. The number q is called the *causal constant* of γ since its sign indicates the causal type of γ. Moreover

$$E := -\langle \gamma', \partial_t \rangle$$

is constant on J since $\partial_t = \partial_{t^*}$ is a Killing vector field. If γ is timelike, E can be interpreted as the *energy* of γ as measured by a Kerr observer (provided the r-coordinate is large enough). The minus sign is a convention which ensures that γ has positive energy if it is future-directed. Furthermore,

$$L := \langle \gamma', \partial_\varphi \rangle$$

is constant on J since $\partial_\varphi = \partial_{\varphi^*}$ is a Killing vector field. If γ is timelike, L can be interpreted as the *angular momentum* of γ as measured by a Kerr observer.
There is one further conserved quantity but it is not induced by a Killing vector field. In order to find it we first generalize the concept of Killing fields.

Definition 4.30. A $(0,k)$-tensor K is called *totally symmetric* if

$$K(X_1, \ldots, X_k) = K(X_{\sigma(1)}, \ldots, X_{\sigma(k)})$$

for any permutation $\sigma \in S_k$.

For $k = 1$ this is a void condition while for $k = 2$ this is the condition of a bilinear form being symmetric.

Definition 4.31. A totally symmetric tensor field K on a semi-Riemannian manifold is called a *Killing tensor field* if

$$(\nabla_X K)(X, \dots, X) = 0$$

for all tangent vectors X.

Example 4.32. **(1)** Let Y be a Killing vector field and let K be the corresponding 1-form, i.e.

$$K(X) = \langle Y, X\rangle$$

for all tangent vectors X. Then K is a Killing tensor field. Indeed, for any vector field X we find

$$\begin{aligned}(\nabla_X K)(X) &= \partial_X(K(X)) - K(\nabla_X X)\\ &= \partial_X\langle Y, X\rangle - \langle Y, \nabla_X X\rangle\\ &= \langle \nabla_X Y, X\rangle = 0.\end{aligned}$$

This example can be generalized, see Exercise 4.21.
(2) Each semi-Riemannian manifold has a canonical Killing $(0, 2)$-tensor field, namely the metric g itself. It is symmetric and is even parallel, $\nabla g = 0$. In particular, it is a Killing tensor field.

Killing tensor fields lead to quantities which are conserved along geodesics. This makes them interesting for us.

Lemma 4.33. *Let K be a Killing tensor field on a semi-Riemannian manifold M and let γ be a geodesic in M. Then the function*

$$s \mapsto K(\gamma'(s), \dots, \gamma'(s))$$

is constant.

Proof. We differentiate:

$$\begin{aligned}&\tfrac{d}{ds}K(\gamma'(s), \dots, \gamma'(s))\\ &\quad = (\nabla_{\gamma'(s)}K)(\gamma'(s), \dots, \gamma'(s)) + K(\tfrac{\nabla}{ds}\gamma'(s), \gamma'(s), \dots, \gamma'(s)) + \ldots + K(\gamma'(s), \dots, \gamma'(s), \tfrac{\nabla}{ds}\gamma'(s))\\ &\quad = 0\end{aligned}$$

where the first term vanishes because K is a Killing tensor field and the others because γ is a geodesic. □

In order to apply this, we construct a Killing $(0, 2)$-tensor field on $\mathrm{Kerr}_{m,a}$. We recall the canonical vector field V defined in Definition 4.19 and we put

$$L_\pm := \tfrac{1}{\Delta}V \pm \partial_r.$$

Using Lemma 4.20 we see that the vector fields L_+ and L_- are lightlike:

$$\langle L_\pm, L_\pm\rangle = \tfrac{1}{\Delta^2}\langle V, V\rangle + g_{rr} = -\tfrac{\rho^2}{\Delta} + g_{rr} = 0.$$

We define a symmetric $(0,2)$-tensor field K on $\mathrm{Kerr}_{m,a}$ by

$$K(X_1, X_2) := \tfrac{\Delta}{2}(\langle L_+, X_1\rangle\langle L_-, X_2\rangle + \langle L_+, X_2\rangle\langle L_-, X_1\rangle) + r^2\langle X_1, X_2\rangle.$$

Lemma 4.34. *The symmetric $(0,2)$-tensor field K on $\mathrm{Kerr}_{m,a}$ is a Killing tensor field.*

Proof. One computes ∇K and checks that for an arbitrary tangent vector X we find $(\nabla_X K)(X, X) = 0$, see the SageMath computation B.2.4. □

Lemma 4.35. *For any $X \in T\mathrm{Kerr}_{m,a}$ we have*

$$\begin{aligned} K(X,X) &= -\frac{\rho^4 dr(X)^2}{\Delta} + \frac{\langle X, V\rangle^2}{\Delta} + r^2\langle X, X\rangle \\ &= \rho^4 d\vartheta(X)^2 + \frac{\langle X, W\rangle^2}{\sin^2\vartheta} - a^2\langle X, X\rangle\cos^2\vartheta. \end{aligned}$$

Proof. This can be checked by direct computation, see the SageMath computation B.2.5. □

Given a geodesic γ we define the auxiliary quantities

$$\begin{aligned} v &:= -\langle\gamma', V\rangle = -\langle\gamma', (r^2+a^2)\partial_t + a\partial_\varphi\rangle = (r^2+a^2)E - aL, \\ w &:= \langle\gamma', W\rangle = \langle\gamma', \partial_\varphi + a\sin^2\vartheta\partial_t\rangle = -a\sin^2\vartheta E + L. \end{aligned}$$

Lemma 4.33 together with Lemma 4.35 with $X = \gamma' = t'\partial_t + r'\partial_r + \vartheta'\partial_\vartheta + \varphi'\partial_\varphi$ yields

Corollary 4.36. *For any geodesic γ in $\mathrm{Kerr}_{m,a}$ the function*

$$-\frac{\rho^4(r')^2}{\Delta} + \frac{v^2}{\Delta} + r^2 q = \rho^4(\vartheta')^2 + \frac{w^2}{\sin^2\vartheta} - a^2 q\cos^2\vartheta$$

is constant. □

Definition 4.37. The constant

$$C := \rho^4(\vartheta')^2 + \frac{w^2}{\sin^2\vartheta} - a^2 q\cos^2\vartheta$$

is called the *Carter constant* of γ.

We can now reduce the geodesic equations in $\mathrm{Kerr}_{m,a}$ to a system of 4 equations of first order.

Theorem 4.38. *Let $\gamma\colon J \to \mathrm{Kerr}_{m,a}$ be a geodesic. Then*

$$\rho^2 t' = aw + \frac{(r^2+a^2)v}{\Delta}, \tag{13}$$

$$\rho^2\varphi' = \frac{w}{\sin^2\vartheta} + \frac{av}{\Delta}, \tag{14}$$

$$\rho^4(r')^2 = (r^2 q - C)\Delta + v^2, \tag{15}$$

$$\rho^4(\vartheta')^2 = C - \frac{w^2}{\sin^2\vartheta} + a^2 q\cos^2\vartheta. \tag{16}$$

Proof. Equations (15) and (16) follow directly from Corollary 4.36 and the definition of the Carter constant.
As to the other two equations, the Boyer-Lindquist identities (BL1)-(BL4) yield

$$\begin{aligned}
\langle\partial_t, V\rangle &= \langle\partial_t, (r^2+a^2)\partial_t + a\partial_\varphi\rangle = (r^2+a^2)g_{tt} + ag_{t\varphi} = -\Delta,\\
\langle\partial_\varphi, V\rangle &= \langle\partial_\varphi, V\rangle = (r^2+a^2)g_{\varphi t} + ag_{\varphi\varphi} = \Delta a\sin^2\vartheta,\\
\langle\partial_t, W\rangle &= \langle\partial_t, \partial_\varphi + a\sin^2\vartheta\partial_t\rangle = g_{t\varphi} + a\sin^2\vartheta g_{tt} = -a\sin^2\vartheta,\\
\langle\partial_\varphi, W\rangle &= \langle\partial_\varphi, \partial_\varphi + a\sin^2\vartheta\partial_t\rangle = g_{\varphi\varphi} + a\sin^2\vartheta g_{t\varphi} = (r^2+a^2)\sin^2\vartheta.
\end{aligned}$$

Thus we obtain

$$\begin{aligned}
v &= -t'\langle\partial_t, V\rangle - \varphi'\langle\partial_\varphi, V\rangle = t'\Delta - \varphi'\Delta a\sin^2\vartheta,\\
w &= t'\langle\partial_t, W\rangle + \varphi'\langle\partial_\varphi, W\rangle = -t'a\sin^2\vartheta + \varphi'(r^2+a^2)\sin^2\vartheta.
\end{aligned}$$

This implies

$$\frac{v}{\Delta} = t' - \varphi' a\sin^2\vartheta, \qquad \frac{w}{a\sin^2\vartheta} = -t' + \varphi'\frac{r^2+a^2}{a}.$$

Solving for φ' and t' we find Equations (13) and (14). □

Remark 4.39. Let $\gamma\colon J \to \mathrm{Kerr}_{m,a}$ be a geodesic.

(1) Let $s_0 \in J$. Then γ is uniquely determined by $\gamma(s_0)$, q, E, L, C, $\operatorname{sign}(r'(s_0))$ and $\operatorname{sign}(\vartheta'(s_0))$ as long as $r'(s) \neq 0$ and $\vartheta'(s) \neq 0$.
(2) Let $\alpha \in \mathbb{R}$ and consider the reparametrized geodesic $\tilde{\gamma}(s) = \gamma(\alpha s)$. Then $\tilde{\gamma}$ has the conserved quantities $\tilde{q} = \alpha^2 q$, $\tilde{E} = \alpha E$, $\tilde{L} = \alpha L$ and $\tilde{C} = \alpha^2 C$.
(3) If γ is causal, then we have $q \leq 0$ and hence $C \geq 0$.
(4) We note that (15) and (16) do not involve functions of t and φ. Thus one can start by solving the system of these two equations first. Then one knows ϱ^2 and the right hand sides of (13) and (14) and one can get t and φ by a simple integration.

In the remainder of this section we restrict our attention for the sake of simplicity to geodesics in the equatorial hyperplane $\mathrm{Eq} = \{(t, r, \sigma) \in \mathrm{Kerr}_{m,a} \,|\, \vartheta(\sigma) = \frac{\pi}{2}\}$. Since Eq is a totally geodesic hypersurface, every geodesic starting in Eq whose initial velocity vector is tangent to Eq will remain in Eq. Since $w = L - a\sin^2\vartheta E$ we get by (16)

$$0 = \varrho^4(\vartheta')^2 = C - w(\pi/2)^2 = C - (L - aE)^2$$

and hence

$$C = (L - aE)^2.$$

Inserting this into (15) and using $v = (r^2 + a^2)E - aL$ we get

$$\begin{aligned} r^4(r')^2 &= (r^2 q - (L - aE)^2)(r^2 + a^2 - 2mr) + ((r^2 + a^2)E - aL)^2 \\ &= (q + E^2)r^4 - 2mqr^3 + (a^2(q + E^2) - L^2)r^2 + 2m(L - aE)^2 r. \end{aligned}$$

Thus we obtain $(r')^2 + U(r) = E^2$ where

$$U(r) := \frac{2mq}{r} + \frac{L^2 - a^2(q + E^2)}{r^2} - \frac{2m(L - aE)^2}{r^3} - q$$

is the effective potential. We have $U(r) \leq E^2$ and thus for given E^2 we can find out the possible values of r using this inequality. Furthermore we have $r' = \pm(E^2 - U(r))^{1/2}$ and therefore

$$r'' = \pm\tfrac{1}{2}(E^2 - U(r))^{-1/2}(-U'(r))r' = -\tfrac{1}{2}U'(r).$$

If for some $r_1 \in \mathbb{R}$ we have $E^2 = U(r_1)$ and $U'(r_1) \neq 0$ then r_1 is a turning point for the radial motion of the geodesic. Namely a geodesic with energy $E^2 = U(r_1)$ reaches $r = r_1$ at a finite parameter value $s_1 \in J$ and we have $r'(s_1) = 0$ and $r''(s_1) \neq 0$, i.e. the radial motion changes its direction.

Let us now discuss the function φ, i.e. the angular motion of the geodesic. We consider lightlike geodesics, i.e. $q = 0$, and we assume that $E > 0$ which for large r means that γ is future-directed. We also assume that $L \neq 0$ and $L \neq aE$.
We start by studying *turning points of the angular motion*. These are points where the geodesic changes its direction of rotation. We define

$$r_w := \frac{2m(L - aE)}{L} \neq 0.$$

We claim that if we are away from the horizons, i.e. $r_w \notin \{r_+, r_-\}$, then the geodesic has a turning point of the angular motion whenever its radial coordinate is $r = r_w$. Furthermore there are no other turning points for the angular motion.
Note that in a Schwarzschild spacetime $r_w = r_+ = 2m$, so Schwarzschild is excluded here. This is not surprising because we already know from the detailed discussion in Section 4.1 that in a Schwarzschild spacetime geodesics do not reverse their angular motion.
In order to understand turning points, we first look for all possible such points by taking $\varphi' = 0$. By (14) we get

$$\begin{aligned} 0 &= \Delta r^2 \varphi' \\ &= \Delta(L - aE) + av \\ &= (r^2 + a^2 - 2mr)(L - aE) + a((r^2 + a^2)E - aL) \\ &= Lr^2 - 2m(L - aE)r. \end{aligned} \tag{17}$$

Since $r \neq 0$ the only possible turning point for the angular motion occurs at $r = r_w$. By (14) we have

$$r^2\varphi' = L - aE + \frac{av(r)}{\Delta(r)} =: f(r).$$

We take the derivative with respect to s and we get

$$r^2\varphi'' + 2rr'\varphi' = \frac{df}{dr}r'.$$

We multiply by r^2 and we obtain

$$r^4\varphi'' = \left(\frac{df}{dr} - \frac{2f(r)}{r}\right)r^2r'.$$

We take the square of both sides and we use equation (15) to get

$$(r^4\varphi'')^2 = \left(\frac{df}{dr} - \frac{2f(r)}{r}\right)^2(-(L - aE)^2\Delta + v^2).$$

At $r = r_w$ we have

$$\begin{aligned} \Delta(r_w) &= \frac{a^2L^2 - 4m^2aE(L - aE)}{L^2} \\ v(r_w)^2 &= (L - aE)^2\left(\frac{16m^4E^2(L - aE)^2}{L^4} + a^2 - \frac{8m^2aE(L - aE)}{L^2}\right) \\ \frac{df}{dr}(r_w) &= \frac{2mL^2(L - aE)}{a^2L^2 - 4m^2aE(L - aE)} \end{aligned}$$

and thus

$$(r^4\varphi'')^2|_{r=r_w} = \frac{16m^4E(L - aE)^5}{a(a^2L^2 - 4m^2aE(L - aE))}.$$

Since

$$(r_w - r_+)(r_w - r_-) = \frac{a^2L^2 - 4m^2aE(L - aE)}{L^2}$$

we get

$$(r^4\varphi'')^2|_{r=r_w} = \frac{16m^4E(L-aE)^5}{aL^2(r_w-r_+)(r_w-r_-)}.$$

Since by hypothesis we have $r_w \notin \{0, r_+, r_-\}$ we get $\varphi'' \neq 0$ at $r = r_w$. Thus at r_w there is the unique point where the geodesic changes its circling direction around the black hole.

Next we examine the *angular motion as the geodesic approaches the horizons* $H_\pm = \{r = r_\pm\}$. We claim that if $r_w \notin \{r_+, r_-\}$ then $\varphi(s)$ is unbounded as $r \to r_\pm$, i.e. the geodesic circles around the horizon infinitely many times.

In order to prove this we first note that we may assume $v(r_+) \neq 0$ and $v(r_-) \neq 0$. Namely, if $v(r_+) = 0$ then since $\Delta(r_+) = 0$ we have $(L-aE)\Delta(r_+) + av(r_+) = 0$ and thus by (17) we get $r_+ = r_w$ contradicting our hypothesis. Thus $v(r_+) \neq 0$ and the same argument shows $v(r_-) \neq 0$. Let $\gamma\colon [s_0, s_1) \to \mathrm{Kerr}_{m,a}$ be a geodesic such that $r(s) \to r_1$ as $s \to s_1$ where $r_1 = r_+$ or $r_1 = r_-$. We consider

$$\lim_{s\to s_1} \varphi(s) - \varphi(s_0) = \int_{s_0}^{s_1} \varphi'(s)ds = \int_{s_0}^{s_1} \Big(L - aE + \frac{av(r)}{\Delta(r)}\Big)\frac{ds}{r^2}$$

where we have used equation (14). For s close to s_1 we have $r(s) \geq c_1 r_1$ for some $c_1 > 0$. Thus we have

$$\Big|\int_{s_0}^{s_1} \frac{L-aE}{r^2}\, ds\Big| < \infty.$$

We take r as the new integration variable, we define $r_0 := r(s_0)$ and we use equation (15) to get

$$\begin{aligned}\int_{s_0}^{s_1} \frac{av(r)}{\Delta(r)}\frac{ds}{r^2} &= a\int_{r_0}^{r_1} \frac{v(r)}{\Delta(r)}\frac{dr}{r^2 r'} = \pm a \int_{r_0}^{r_1} \frac{v}{\Delta}(v^2 - (L-aE)^2\Delta)^{-1/2}\, dr \\ &= \pm a \int_{r_0}^{r_1} \frac{1}{(r-r_+)(r-r_-)} \underbrace{(1-(L-aE)^2 v^{-2}\Delta)^{-1/2}}_{=:D}\, dr.\end{aligned}$$

Since $v(r_1) \neq 0$ and $\Delta(r_1) = 0$ we have $D \to 1$ as $s \to s_1$. It follows that there exists $c_2 > 0$ such that

$$\Big|\int_{s_0}^{s_1} \frac{av(r)}{\Delta(r)}\frac{ds}{r^2}\Big| \geq c_2 \Big|\int_{r_0}^{r_1} \frac{dr}{(r-r_+)(r-r_-)}\Big|.$$

If $r_+ \neq r_-$ there is $c_3 > 0$ such that $\frac{1}{(r-r_-)(r-r_+)} \geq c_3 \frac{1}{r-r_1}$. Thus in both cases $r_+ = r_-$ and $r_+ \neq r_-$ this integral does not converge since the integrals

$$\int_{r_0}^{r_1} \frac{dr}{r-r_1}, \quad \int_{r_0}^{r_1} \frac{dr}{(r-r_1)^2}$$

do not converge. Thus $\varphi(s)$ is unbounded as $s \to s_1$.

Next we consider geodesics approaching the ring singularity $\Sigma = \{r = 0\} \cap \mathrm{Eq}$. We claim that $\varphi(s)$ converges as $r \to 0$. In order to prove this we write as above

$$\lim_{s\to s_1} \varphi(s) - \varphi(s_0) = \int_{s_0}^{s_1} \varphi'(s)ds = \int_{s_0}^{s_1} \Big(L - aE + \frac{av(r)}{\Delta(r)}\Big)\frac{ds}{r^2}$$

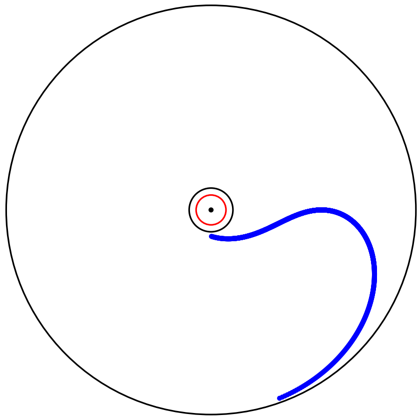


Figure 88. In this and the following orbit plots, the radial coordinate r is drawn in an exponential scale to accommodate negative values of r, too. In particular, the ring singularity Σ corresponds to the red circle of radius 1 in these diagrams. This plot shows a lightlike geodesic in Boyer-Lindquist block II spiraling to the inner and outer horizons drawn in black. Note the reversal of circling direction. The values are $m = 1.5$, $a = 1$, $L = 6$, and $E = 2$.

where we have used equation (14). As above we take r as the new integration variable, we define $r_0 := r(s_0)$ and we use equation (15) to get

$$\int_{s_0}^{s_1} \Big(L - aE + \frac{av(r)}{\Delta(r)}\Big)\frac{ds}{r^2} = \pm \int_{r_0}^{0} \Big(L - aE + \frac{av(r)}{\Delta(r)}\Big)(v(r)^2 - (L - aE)^2\Delta(r))^{-1/2}\, dr.$$

We have

$$L - aE + \frac{av(r)}{\Delta(r)} = \frac{Lr^2 - 2m(L - aE)r}{r^2 + a^2 - 2mr}$$

and

$$(v(r)^2 - (L - aE)^2\Delta(r))^{-1/2} = (E^2r^4 - (L - aE)(L + aE)r^2 + 2mr(L - aE)^2)^{-1/2}.$$

Thus the integrand is asymptotic to $r^{1/2}$ as $r \to 0$ and the integral converges.

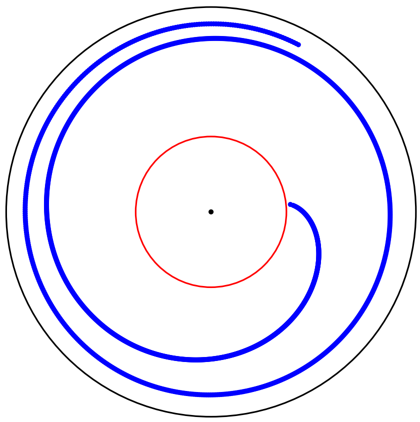


Figure 89. Lightlike geodesic falling into the ring singularity (red) and spiraling to the inner horizon (black). The values are $m = a = 1$, $L = 2$, and $E = 1$.

Next we discuss the function r, i.e. the radial motion of the geodesic. As before, we consider lightlike geodesics, i.e. $q = 0$, and we assume that $E > 0$. The effective potential defined above becomes

$$U(r) = \frac{L^2 - a^2E^2}{r^2} - \frac{2m(L-aE)^2}{r^3}.$$

If $L \neq aE$ we have

$$\lim_{r\searrow 0} U(r) = -\infty, \quad \lim_{r\nearrow 0} U(r) = \infty, \quad \lim_{r\to\pm\infty} U(r) = 0.$$

If $L \notin \{\pm aE\}$ we have

$$U(r) = 0 \Leftrightarrow r(L^2 - a^2E^2) = 2m(L-aE)^2 \Leftrightarrow r = \frac{2m(L-aE)}{L+aE}$$

and

$$\begin{aligned}
\frac{dU}{dr} = 0 &\Leftrightarrow -2\frac{L^2-a^2E^2}{r^3} + 3\frac{2m(L-aE)^2}{r^4} = 0 \\
&\Leftrightarrow 2(L^2-a^2E^2)r = 6m(L-aE)^2 \\
&\Leftrightarrow r = \frac{3m(L-aE)}{L+aE} =: r_{\text{ext}}.
\end{aligned}$$

We compute

$$\begin{aligned}
U(r_{\text{ext}}) &= \frac{(L^2-a^2E^2)(L+aE)^2}{(3m(L-aE))^2} - \frac{2m(L-aE)^2(L+aE)^3}{(3m(L-aE))^3} \\
&= \frac{(L+aE)^3}{9m^2(L-aE)} - \frac{2(L+aE)^3}{27m^2(L-aE)} \\
&= \frac{1}{27m^2}\frac{(L+aE)^3}{L-aE}.
\end{aligned}$$

(1) *Case* $L > aE$. Then $r_{\text{ext}} > 0$, $U(r_{\text{ext}}) > 0$ and r_{ext} is a local maximum of the function U.
(a) *Subcase* $E^2 > U(r_{\text{ext}}) = \frac{1}{27m^2}\frac{(L+aE)^3}{L-aE}$.

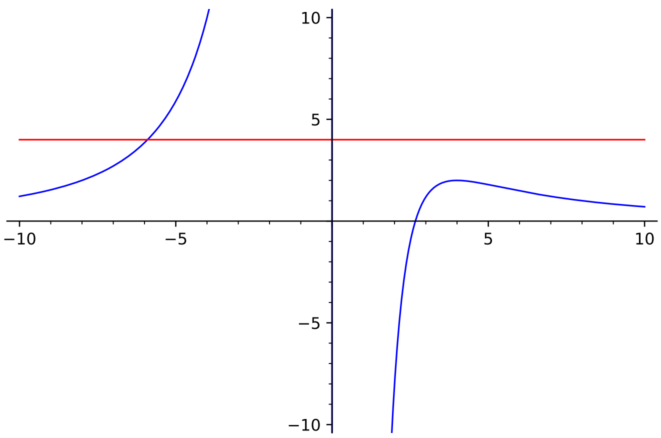


Figure 90. The effective potential $U(r)$ for $m = 2$, $a = 1$, $L = 10$, and $E = 2$. The red line shows the energy threshold at E^2.

For $r > 0$ we have

$$(r')^2 = E^2 - U(r) \geq E^2 - U(r_{\text{ext}}) > 0$$

and thus $|r'(s)| \geq (E^2 - U(r_{\text{ext}}))^{1/2}$ for all s where the right hand side is a constant independent of s. Thus ingoing geodesics hit the ring singularity at a finite value of s and outgoing geodesics escape to infinity. We call this motion a collision-escape orbit.
As to negative r, there is a unique value $r_1 < 0$ such that $U(r_1) = E^2$. Moreover $U'(r_1) \neq 0$, i.e. r_1 is a turning point for the radial motion. Thus for $r < 0$ the function r cannot take values in $(r_1, 0)$ and geodesics hitting $r = r_1$ turn around and escape to $-\infty$. We call this motion a fly-by orbit.
(b) *Subcase* $E^2 = U(r_{\text{ext}})$.

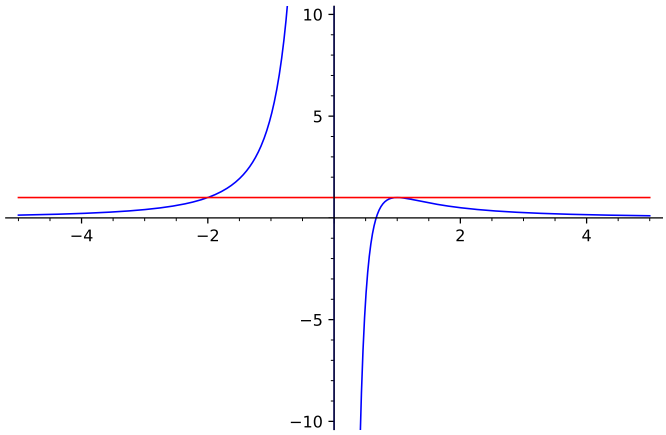


Figure 91. The effective potential $U(r)$ for $m = a = 1$, $L = 2$, and $E = 1$. The red line shows the energy threshold at E^2.

(i) $r(0) > r_{\text{ext}}$: Outgoing geodesics escape to infinity, ingoing ones satisfy $r(s) \to r_{\text{ext}}$ as $s \to \infty$.

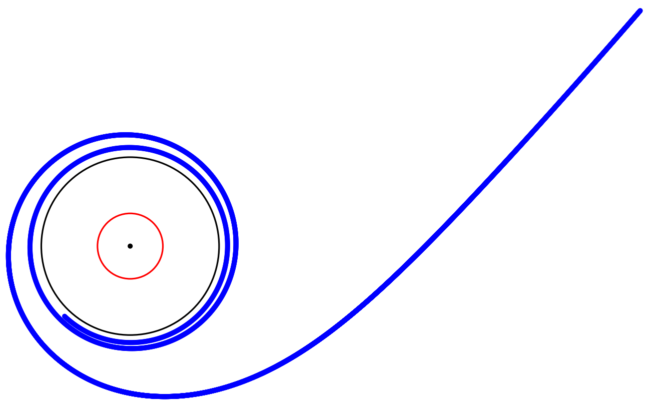


Figure 92. Lightlike geodesic spiraling to the horizon and escaping to infinity. The values are $m = a = 1$, $L = 2$, and $E = 1$.

(ii) $r(0) = r_{\text{ext}}$: We see that the functions $\vartheta \equiv \frac{\pi}{2}$, $r \equiv r_{\text{ext}}$ and

$$t(s) = t_0 + s\frac{1}{r_{\text{ext}}^2}\Big(a(L - aE) + \frac{(r_{\text{ext}}^2 + a^2)v(r_{\text{ext}})}{\Delta(r_{\text{ext}})}\Big)$$
$$\varphi(s) = \varphi_0 + s\frac{1}{r_{\text{ext}}^2}\Big(L - aE + \frac{av(r_{\text{ext}})}{\Delta(r_{\text{ext}})}\Big)$$

solve the equations (13)-(16) provided that $r_{\text{ext}} \notin \{r_+, r_-\}$. This geodesic describes a photon circling around the black hole with constant angular velocity and at a constant distance. Note that for fixed m and a the distance $r_{\text{ext}} = \frac{3m(L-aE)}{L+aE}$ can take all values in $(0, 3m)$ for suitable choices of L and E satisfying $L > aE$. Hence, there is a whole photon annular region in contrast to the Schwarzschild model with its photon sphere.

(iii) $0 < r(0) < r_{\text{ext}}$: Outgoing geodesics satisfy $r(s) \to r_{\text{ext}}$ as $s \to \infty$, ingoing geodesics hit the ring singularity.

(iv) $r(0) < 0$: The discussion for negative r is as in case a) and leads to a fly-by orbit.

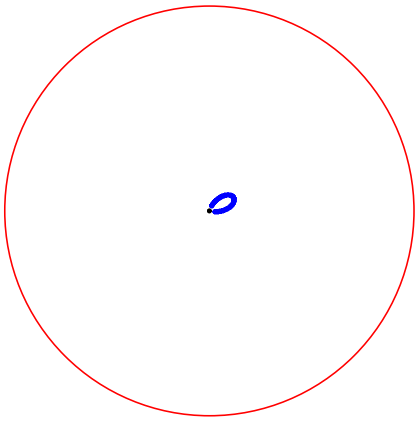


Figure 93. Lightlike fly-by orbit in block III. The values are $m = a = 1$, $L = 2$, and $E = 1$. Note that the origin corresponds to $r = -\infty$.

(c) *Subcase* $E^2 < U(r_{\text{ext}})$.

(i) $r(0) > r_{\text{ext}}$: fly-by orbit

(ii) $0 < r(0) < r_{\text{ext}}$: collision orbit

(iii) $r(0) < 0$: fly-by orbit

(2) *Case* $L = aE$: Then $U \equiv 0$ and thus $|r'|$ is constant. In both cases $r(0) < 0$ and $r(0) > 0$ we obtain a collision-escape orbit.

(3) *Case* $-aE < L < aE$: Then $r_{\text{ext}} < 0$, $U(r_{\text{ext}}) < 0$ and r_{ext} is a local minimum of the function U. For $r > 0$ we have $(r')^2 = E^2 - U(r) \geq E^2$ and thus $|r'(s)| \geq E$ independently of s. We obtain a collision-escape orbit. There is exactly one r_1 such that $U(r_1) = E^2$ and we have $r_1 < 0$. Thus the function r cannot take values in $(r_1, 0)$ and for $r(0) \leq r_1$ we get a fly-by orbit.

(4) *Case* $L = -aE$: Then $U(r) = -\frac{8ma^2E^2}{r^3}$ has no local extrema. Apart from that the discussion is analogous to the case $-aE < L < aE$.

(5) *Case* $L < -aE$: Then $r_{\text{ext}} > 0$, $U(r_{\text{ext}}) > 0$ and r_{ext} is a local maximum of the function U. The discussion is analogous to the case $L > aE$.

The discussion of geodesics in many other cases can be found in O'Neill's book [8].

4.2.8. The Kerr metric is a vacuum solution

The Kerr metric $g_{(m,a)}$ is indeed a solution to the vacuum Einstein field equations.

Proposition 4.40. *The Ricci curvature of the Kerr metric $g_{(m,a)}$ vanishes.*

The proof is a straighforward but lengthy and tedious computation. Fortunately, it can be delegated to SageMath, see Section B.2.6.

4.3. Exercises

4.1. Let M be the Schwarzschild spacetime with mass $m \geq 0$. At each point p let $Z \subset T_pM$ be the plane spanned by $\frac{\partial}{\partial t}$ and $\frac{\partial}{\partial r}$. Show that the sectional curvature of a plane $E \subset T_pM$ satisfies:

$$K(E) = \begin{cases} \frac{2m}{r^3}, & \text{if } E \text{ is tangential to } S^2 \text{ or if } E = Z; \\ -\frac{m}{r^3}, & \text{if } E \text{ is spanned by a vector tangential to } S^2 \text{ and by one in } Z. \end{cases}$$

4.2. In a Schwarzschild spacetime with $m > 0$ there are photons circling at constant Schwarzschild distance $r = 3m$ around the black hole. Compute the corresponding lightlike geodesics.

4.3. Let γ be the wordline of a freely falling material particle in Schwarzschild spacetime M with $m > 0$ which falls into the horizon from the outside. In other words, $\gamma : [0, T) \to M$ is a timelike geodesic with $\gamma(s) = (t(s), r(s), \theta(s), \varphi(s))$ where $r(s) > 2m$ and $\lim_{s\to T} r(s) = 2m$. Show:
(a) $T < \infty$, i.e. the particle reaches the horizon in finite proper time.
(b) $\lim_{s\to T} t(s) = \infty$, i.e. measured in Schwarzschild time, the particle needs infinite time to reach the horizon.

4.4. Fix $r_0 > 2m > 0$ and consider the hypersurface $\mathscr{H}_{r_0} = \{r = r_0\}$ in the Schwarzschild spacetime for mass m. Show that $\mathscr{H}_{r_0}$ is totally umbilic if and only if $r_0 = 3m$.
Here *totally umbilic* means that the Weingarten map of the hypersurface is a multiple of the identity, i.e. $\nabla_X \nu = \lambda X$ for all X tangent to the hypersurface where λ is a constant and ν is a unit normal field along the hypersurface.

4.5. Compute the Schwarzschild radius of the earth (mass $5.9723 \cdot 10^{24}$ kg), of the moon ($7.349 \cdot 10^{22}$ kg) and of the sun ($1.9884 \cdot 10^{30}$ kg).

4.6. (a) Show that the projection of any causal curve in $\mathsf{Kr} \times S^2$ to Kr is causal in Kr.
(b) Show that every causal curve in $\mathsf{Kr} \times S^2$ can pass the event horizon at most twice.

4.7. (a) Show that for each $s \in \mathbb{R}$ the map $\Phi_s : \mathsf{Kr} \to \mathsf{Kr}$, $(u, v) \mapsto (e^{-s}u, e^s v)$, is an isometry of the Kruskal plane.
(b) Compute the corresponding Killing vector field and express it in the Schwarzschild frame

$\frac{\partial}{\partial t}$ and $\frac{\partial}{\partial r}$.

4.8. Show that in Boyer-Lindquist coordinates the Kerr metric can be written as

$$g_{(m,a)} = \frac{\rho^2}{\Delta} dr \otimes dr + \rho^2 g_{S^2} + a^2 \sin^4\theta \, d\phi \otimes d\phi - dt \otimes dt + \frac{2mr}{\rho^2}(dt - a\sin^2\theta \, d\phi) \otimes (dt - a\sin^2\theta \, d\phi)$$

on $\mathbb{R} \times \mathbb{R} \times S^2 \setminus (\Sigma \cup H \cup A)$.

4.9. Let $\Phi, \Psi : \mathbb{R}^2 \times S^2 \setminus (\Sigma \cup H) \to \mathbb{R}^2 \times S^2 \setminus (\Sigma \cup H)$ be given by

$$\Phi(t, r, \sigma) = (-t, r, \sigma),$$
$$\Psi(t, r, \sigma) = (t, r, \mathrm{S}(\sigma)),$$

where $\mathrm{S} : S^2 \to S^2$ denotes the reflection about a plane which contains the north and the south poles $(0, 0, \pm 1)$. Show:

$$\Phi^* g_{(m,a)} = \Psi^* g_{(m,a)} = g_{(m,-a)},$$

where we now allow all $a \in \mathbb{R}$.

4.10. Let $\alpha : (s_0, s_1) \to I$ be a future-directed causal curve which is contained in A and which falls into the outer horizon, more precisely, $r(\alpha(s)) \to r_+$ and $(r \circ \alpha)'(s) \to r^* \neq 0$ for $s \nearrow s_1$. Show that $(t \circ \alpha)' \to \infty$ as $s \nearrow s_1$.

4.11. We know that for the slowly rotating Kerr solution, the region I does not contain any closed causal curves but III does. Does the rapidly rotating Kerr solution have closed causal curves?

4.12. Let $a < m$ and let $\alpha : J \to I$ be a future-directed timelike curve, parametrized by proper time. Show that

$$(t \circ \alpha)'(s) \geq \left(1 - \frac{2m^2}{r(\alpha(s))^2 + a^2}\right)^{-1/2}$$

for all $s \in J$.

4.13. Show that the Carter time machine C is connected.
Hint: Show that the polynomial in r which characterizes C along a ray $\vartheta \equiv \vartheta_0$ has only one real root for any ϑ_0.

4.14. Express the coordinate vector fields for the Kerr coordinates by those for the Boyer-Lindquist coordinates and vice versa.

4.15. Let $\alpha : (-1, 1) \to \mathrm{Kerr}^*_{m,a}$ be a smooth causal curve with $\alpha(0) \in H$ and $\alpha'(0) = V(\alpha(0))$. Show:

$$\frac{d}{ds}(r^* \circ \alpha)\Big|_{s=0} = \frac{d^2}{ds^2}(r^* \circ \alpha)\Big|_{s=0} = 0.$$

4.16. Let $p \in \mathrm{Kerr}^*_{(m,a)}$. Show that $p \in I$ if and only if there exists a future-directed timelike curve $\alpha : [0, \infty) \to \mathrm{Kerr}^*_{(m,a)}$ with $\alpha(0) = p$ and $\lim_{s\to\infty} r(\alpha(s)) = \infty$.

4.17. Show that the r-coordinate lines (w.r.t. Boyer-Lindquist coordinates) in the equatorial hypersurface Eq are pregeodesics, both as curves in Eq and as curves in $\mathrm{Kerr}_{m,a}$.

4.18. Let $a < m$ and let X be the vector field $X = a\partial_t + \beta\partial_\varphi$ on block I for some constant β. Let $r_0 > r_+$.
(a) Show $r_0^3 > ma^2$.
(b) Determine β such that the integral curves of X at $r = r_0$ and $\vartheta = \frac{\pi}{2}$ are geodesics.
(c) Show that X is timelike for these β and for sufficiently large r_0.
Remark: We have found world lines of massive particles which circle at distance r_0 around the black hole.

4.19. Let M be a semi-Riemannian manifold and let $\Psi : [0, \varepsilon_0) \times M \to M$ be a smooth map such that each $\Psi_\varepsilon := \Psi(\varepsilon, \cdot) : M \to M$ is an isometry and $\Psi_0 = \mathrm{id}$. Let $X := \frac{d}{d\varepsilon}\Psi_\varepsilon|_{\varepsilon=0}$ be the corresponding Killing vector field.
Show that X is a Jacobi field along each geodesic.

4.20. Show that the Kerr-Metrik $g_{m,a}$ cannot be extended across the ring singularity Σ.
Hint: Use SageMath to compute the Kretschmann scalar curvature.

4.21. Let $Y_1, \ldots, Y_k$ be Killing vector fields on a semi-Riemannian manifold. Show that

$$K(X_1, \ldots, X_k) = \sum_{\sigma \in S_k} \langle Y_1, X_{\sigma(1)}\rangle \cdots \langle Y_k, X_{\sigma(k)}\rangle$$

defines a Killing tensor field.

5. Gravitational waves

We want to consider small perturbations of a given solution to the Einstein field equations. These fluctuations are described well by solutions of the linearized field equations. In order to linearize the field equations at a given solution we consider the first variation of the relevant geometric quantities.

5.1. First variation of geometric quantities

Let (M, g) be a semi-Riemannian manifold and let $(g_s)_{s\in(-\varepsilon,\varepsilon)}$ be a smooth 1-parameter family of metrics with $g_0 = g$. Then $h := \frac{\partial}{\partial s}|_{s=0} g_s$ is a $(0, 2)$-tensor field on M. Let ∇^s be the Levi-Civita connection of g_s. Then $\nabla^s - \nabla^0$ is a $(1, 2)$-tensor field on M and thus

$$\lim_{s\to 0} \frac{1}{s}(\nabla^s - \nabla^0) =: \nabla'$$

is also a $(1, 2)$-tensor field on M. We denote by $\nabla := \nabla^0$ the Levi-Civita connection of g. Then ∇' is characterized by the following formula.

Lemma 5.1. *For all $X, Y, Z \in T_pM$, $p \in M$, we have*

$$g(\nabla'(X, Y), Z) = \tfrac{1}{2}\{(\nabla_X h)(Y, Z) + (\nabla_Y h)(X, Z) - (\nabla_Z h)(X, Y)\}.$$

Proof. We differentiate the Koszul formula

$$\begin{aligned} g_s(\nabla^s_X Y, Z) =& \tfrac{1}{2}\{\partial_X g_s(Y, Z) + \partial_Y g_s(X, Z) - \partial_Z g_s(X, Y) \\ &- g_s(X, [Y, Z]) + g_s(Y, [Z, X]) + g_s(Z, [X, Y])\} \end{aligned}$$

with respect to s at $s = 0$ and we get

$$\begin{aligned} h(\nabla_X Y, Z) + g(\nabla'(X, Y), Z) =& \tfrac{1}{2}\{\partial_X h(Y, Z) + \partial_Y h(X, Z) - \partial_Z h(X, Y) \\ &- h(X, [Y, Z]) + h(Y, [Z, X]) + h(Z, [X, Y])\} \\ =& \tfrac{1}{2}\{(\nabla_X h)(Y, Z) + h(\nabla_X Y, Z) + h(Y, \nabla_X Z) \\ &+ (\nabla_Y h)(X, Z) + h(\nabla_Y X, Z) + h(X, \nabla_Y Z) \\ &- (\nabla_Z h)(X, Y) - h(\nabla_Z X, Y) - h(X, \nabla_Z Y) \\ &- h(X, \nabla_Y Z - \nabla_Z Y) + h(Y, \nabla_Z X - \nabla_X Z) + h(Z, \nabla_X Y - \nabla_Y X)\} \\ =& h(\nabla_X Y, Z) + \tfrac{1}{2}\{(\nabla_X h)(Y, Z) + (\nabla_Y h)(X, Z) - (\nabla_Z h)(X, Y)\}. \end{aligned}$$

We finish the proof by subtracting $h(\nabla_X Y, Z)$ on both sides. □

The Riemann curvature tensor of g_s is a $(1,3)$-tensor field on M. Then $R' := \frac{\partial}{\partial s}|_{s=0} R^s$ is a $(1,3)$-tensor field on M and we have the following formula.

Lemma 5.2. *For all X, Y, Z in T_pM, $p \in M$, we have*

$$R'(X,Y)Z = (\nabla_X \nabla')(Y,Z) - (\nabla_Y \nabla')(X,Z).$$

Proof. We differentiate the equation

$$R^s(X,Y)Z = \nabla^s_X \nabla^s_Y Z - \nabla^s_Y \nabla^s_X Z - \nabla^s_{[X,Y]} Z$$

with respect to s at $s = 0$ and we get

$$\begin{aligned} R'(X,Y)Z =& \nabla'(X, \nabla_Y Z) + \nabla_X(\nabla'(Y,Z)) - \nabla'(Y, \nabla_X Z) - \nabla_Y(\nabla'(X,Z)) - \nabla'([X,Y],Z) \\ =& \nabla'(X, \nabla_Y Z) + (\nabla_X \nabla')(Y,Z) + \nabla'(\nabla_X Y, Z) + \nabla'(Y, \nabla_X Z) \\ & - \nabla'(Y, \nabla_X Z) - (\nabla_Y \nabla')(X,Z) - \nabla'(\nabla_Y X, Z) - \nabla'(X, \nabla_Y Z) - \nabla'(\nabla_X Y - \nabla_Y X, Z) \\ =& (\nabla_X \nabla')(Y,Z) - (\nabla_Y \nabla')(X,Z). \end{aligned}$$

□

Before we calculate the variation of the Ricci tensor we define several operators. Let $(E_i)_{i=1}^n$ be a local orthonormal basis of TM and let $\varepsilon_i := g(E_i, E_i)$, $i = 1, ..., n$. We denote by $\Gamma(\odot^2 T^*M)$ the space of all symmetric $(0,2)$-tensor fields on M. We define the *connection Laplacian* acting on $\Gamma(\odot^2 T^*M)$ by

$$\nabla^*\nabla h := -\sum_{i=1}^n \varepsilon_i \nabla^2_{E_i,E_i} h = -\sum_{i=1}^n \varepsilon_i \nabla_{E_i} \nabla_{E_i} h + \sum_{i=1}^n \varepsilon_i \nabla_{\nabla_{E_i} E_i} h.$$

We note that on a Riemannian manifold this operator is elliptic and on a Lorentzian manifold it is hyperbolic. For $h \in \Gamma(\odot^2 T^*M)$ we define

$$(\mathring{R}h)(X,Y) := \sum_{i=1}^n \varepsilon_i h(R(E_i, X)Y, E_i).$$

Then we have $\mathring{R}h \in \Gamma(\odot^2 T^*M)$. For $h_1, h_2 \in \Gamma(\odot^2 T^*M)$ we define

$$(h_1 \circ h_2)(X,Y) := \sum_{i=1}^n \varepsilon_i h_1(X, E_i) h_2(Y, E_i).$$

Then we have $h_1 \circ h_2 + h_2 \circ h_1 \in \Gamma(\odot^2 T^*M)$. We define the *Lichnerowicz operator* acting on $\Gamma(\odot^2 T^*M)$ by

$$\Delta_L h := \nabla^*\nabla h + \mathrm{ric} \circ h + h \circ \mathrm{ric} - 2\mathring{R}h.$$

Finally we define the *divergence* of $h \in \Gamma(\odot^2 T^* M)$ by

$$\operatorname{div} h := \sum_{i=1}^{n} \varepsilon_i (\nabla_{E_i} h)(E_i, \cdot).$$

Note that $\operatorname{div} h$ is a 1-form on M. For 1-forms ω on M we define $\operatorname{div}^* \omega$ by

$$\operatorname{div}^* \omega(X, Y) := -\tfrac{1}{2}\{(\nabla_X \omega)(Y) + (\nabla_Y \omega)(X)\}.$$

Then we have $\operatorname{div}^* \omega \in \Gamma(\odot^2 T^* M)$ and we note that div^* is the L^2-adjoint of div. We define the g-trace of $h \in \Gamma(\odot^2 T^* M)$ by

$$\langle g, h \rangle := \sum_{i=1}^{n} \varepsilon_i h(E_i, E_i).$$

Then $\langle g, h \rangle$ is a smooth function on M. Finally the *Hessian* of a smooth function f on M is defined by

$$\nabla df(X, Y) := (\nabla_X df)(Y).$$

Then we have $\nabla df \in \Gamma(\odot^2 T^* M)$. Let ric^s be the Ricci curvature tensor of g_s considered as a $(0, 2)$-tensor field on M and define $\operatorname{ric}' := \frac{\partial}{\partial s}|_{s=0} \operatorname{ric}^s$. We obtain the following formula.

Lemma 5.3. *We have*

$$\operatorname{ric}' = \tfrac{1}{2}\Delta_L h - \operatorname{div}^* \operatorname{div} h - \tfrac{1}{2}\nabla d \langle g, h \rangle .$$

We note that every term on the right hand side contains second order derivatives of h. In fact, ric' is not elliptic in the Riemannian case and not hyperbolic in the Lorentzian case.

Proof. We will write $\langle \cdot, \cdot \rangle$ instead of $g(\cdot, \cdot)$. By definition $\operatorname{ric}^s(X, Y)$ is the trace of the endomorphism

$$\operatorname{ric}^s(X, Y) = \operatorname{tr}(R^s(\cdot, X)Y).$$

Since the trace is a linear map independent of s, we get $\operatorname{ric}'(X, Y) = \operatorname{tr}(R'(\cdot, X)Y)$. We fix a point $p \in M$ and we extend the vectors $X, Y \in T_p M$ in such a way that $\nabla X|_p = 0$ and $\nabla Y|_p = 0$ at p. Furthermore we can choose the local orthonormal basis in such a way that $\nabla E_i|_p = 0$ at p. Note however that the second derivatives of X, Y, and E_i will not vanish at p in general. Then at the point p we get using Lemmas 5.1, 5.2

$$\begin{aligned}
\mathrm{ric}'(X,Y) &= \sum_{i=1}^{n} \varepsilon_i \langle R'(E_i,X)Y, E_i\rangle \\
&= \sum_{i=1}^{n} \varepsilon_i \{\langle(\nabla_{E_i}\nabla')(X,Y),E_i\rangle - \langle(\nabla_X\nabla')(E_i,Y),E_i\rangle\} \\
&= \sum_{i=1}^{n} \varepsilon_i \{\partial_{E_i}\langle\nabla'(X,Y),E_i\rangle - \partial_X\langle\nabla'(E_i,Y),E_i\rangle\} \\
&= \frac{1}{2}\sum_{i=1}^{n} \varepsilon_i \{\partial_{E_i}((\nabla_X h)(Y,E_i) + (\nabla_Y h)(X,E_i) - (\nabla_{E_i}h)(X,Y)) \\
&\quad - \partial_X((\nabla_{E_i}h)(Y,E_i) + (\nabla_Y h)(E_i,E_i) - (\nabla_{E_i}h)(E_i,Y))\} \\
&= \frac{1}{2}\sum_{i=1}^{n} \varepsilon_i \{\partial_{E_i}((\nabla_X h)(Y,E_i) + (\nabla_Y h)(X,E_i) - (\nabla_{E_i}h)(X,Y)) - \partial_X((\nabla_Y h)(E_i,E_i))\}. \quad (1)
\end{aligned}$$

Next we analyze each term on the right hand side. We have

$$\begin{aligned}
\sum_{i=1}^{n} \varepsilon_i \partial_X((\nabla_Y h)(E_i,E_i)) &= \sum_{i=1}^{n} \varepsilon_i \partial_X(\partial_Y h(E_i,E_i) - 2h(\nabla_Y E_i, E_i)) \\
&= \partial_X\partial_Y\langle g,h\rangle - 2\sum_{i=1}^{n}\varepsilon_i h(\nabla_X\nabla_Y E_i, E_i).
\end{aligned}$$

The second term on the right hand side of this equation vanishes. Namely we have

$$\sum_{i=1}^{n} \varepsilon_i h(\nabla_X\nabla_Y E_i, E_i) = \sum_{i,j=1}^{n} \varepsilon_i\varepsilon_j \left\langle \nabla_X\nabla_Y E_i, E_j\right\rangle h(E_j,E_i)$$

and using that $\left\langle E_i, E_j\right\rangle$ is constant we get

$$\begin{aligned}
\left\langle \nabla_X\nabla_Y E_i, E_j\right\rangle &= \partial_X\left\langle \nabla_Y E_i, E_j\right\rangle = \partial_X(\partial_Y\left\langle E_i,E_j\right\rangle - \left\langle E_i, \nabla_Y E_j\right\rangle) = -\partial_X\left\langle E_i,\nabla_Y E_j\right\rangle \\
&= -\left\langle E_i, \nabla_X\nabla_Y E_j\right\rangle,
\end{aligned}$$

i.e. $\left\langle \nabla_X\nabla_Y E_i, E_j\right\rangle$ is antisymmetric in i, j, while $\varepsilon_i\varepsilon_j h(E_j,E_i)$ is symmetric in i, j. Using this fact and renaming i and j we get

$$\begin{aligned}
\sum_{i,j=1}^{n} \varepsilon_i\varepsilon_j \left\langle \nabla_X\nabla_Y E_i, E_j\right\rangle h(E_j,E_i) &= \sum_{i,j=1}^{n} \varepsilon_j\varepsilon_i \left\langle \nabla_X\nabla_Y E_j, E_i\right\rangle h(E_i,E_j) \\
&= -\sum_{i,j=1}^{n} \varepsilon_i\varepsilon_j \left\langle \nabla_X\nabla_Y E_i, E_j\right\rangle h(E_j,E_i)
\end{aligned}$$

and therefore this term vanishes. Altogether we obtain

$$\sum_{i=1}^{n} \varepsilon_i \partial_X((\nabla_Y h)(E_i, E_i)) = \partial_X \partial_Y \langle g, h \rangle = \nabla d \langle g, h \rangle (X, Y). \tag{2}$$

Next we compute

$$\begin{aligned}
\sum_{i=1}^{n} \varepsilon_i \partial_{E_i}((\nabla_X h)(Y, E_i)) = \sum_{i=1}^{n} \varepsilon_i \partial_{E_i}(\partial_X h(Y, E_i) - h(\nabla_X Y, E_i) - h(Y, \nabla_X E_i))& \\
= \partial_X \sum_{i=1}^{n} \varepsilon_i \partial_{E_i} h(Y, E_i) - \sum_{i=1}^{n} \varepsilon_i h(\nabla_{E_i} \nabla_X Y, E_i) - \sum_{i=1}^{n} \varepsilon_i h(Y, \nabla_{E_i} \nabla_X E_i)& \\
= \partial_X \Big(\operatorname{div} h(Y) + \sum_{i=1}^{n} \varepsilon_i h(\nabla_{E_i} Y, E_i) + \sum_{i=1}^{n} \varepsilon_i h(Y, \nabla_{E_i} E_i)\Big)& \\
- \sum_{i=1}^{n} \varepsilon_i h(\nabla_{E_i} \nabla_X Y, E_i) - \sum_{i=1}^{n} \varepsilon_i h(Y, \nabla_{E_i} \nabla_X E_i)& \\
= (\nabla_X \operatorname{div} h)(Y) + \sum_{i=1}^{n} \varepsilon_i h(\nabla_X \nabla_{E_i} Y, E_i) + \sum_{i=1}^{n} \varepsilon_i h(Y, \nabla_X \nabla_{E_i} E_i)& \\
- \sum_{i=1}^{n} \varepsilon_i h(\nabla_{E_i} \nabla_X Y, E_i) - \sum_{i=1}^{n} \varepsilon_i h(Y, \nabla_{E_i} \nabla_X E_i)& \\
= (\nabla_X \operatorname{div} h)(Y) + \sum_{i=1}^{n} \varepsilon_i h(R(X, E_i)Y, E_i) + \sum_{i=1}^{n} \varepsilon_i h(Y, R(X, E_i)E_i)& \\
= (\nabla_X \operatorname{div} h)(Y) - \sum_{i=1}^{n} \varepsilon_i h(R(E_i, X)Y, E_i)& \\
+ \sum_{i,j=1}^{n} \varepsilon_i \varepsilon_j \Big\langle R(X, E_i)E_i, E_j \Big\rangle h(Y, E_j)& \\
= (\nabla_X \operatorname{div} h)(Y) - (\mathring{R} h)(X, Y) + (\operatorname{ric} \circ h)(X, Y).&
\end{aligned} \tag{3}$$

Analogously we get

$$\sum_{i=1}^{n} \varepsilon_i \partial_{E_i}((\nabla_X h)(Y, E_i)) = (\nabla_Y \operatorname{div} h)(X) - (\mathring{R} h)(X, Y) + (h \circ \operatorname{ric})(X, Y). \tag{4}$$

We also compute

$$\sum_{i=1}^{n} \varepsilon_i \partial_{E_i}((\nabla_{E_i} h)(X, Y)) = \sum_{i=1}^{n} \varepsilon_i (\nabla_{E_i} \nabla_{E_i} h)(X, Y) = -\nabla^* \nabla h(X, Y). \tag{5}$$

Inserting (2), (3), (4), (5) into (1) we obtain

$$\begin{aligned}
\operatorname{ric}'(X, Y) &= -\tfrac{1}{2} \nabla d \langle g, h \rangle (X, Y) + \tfrac{1}{2} \nabla^* \nabla h(X, Y) + \tfrac{1}{2}((\nabla_X \operatorname{div} h)(Y) + (\nabla_Y \operatorname{div} h)(X)) \\
&\quad - \mathring{R} h(X, Y) + \tfrac{1}{2}(\operatorname{ric} \circ h)(X, Y) + \tfrac{1}{2}(h \circ \operatorname{ric})(X, Y) \\
&= -\tfrac{1}{2} \nabla d \langle g, h \rangle (X, Y) - (\operatorname{div}^* \operatorname{div} h)(X, Y) + \tfrac{1}{2} \Delta_L h(X, Y).
\end{aligned}$$

which concludes the proof. □

We note that the operator ric' is not elliptic for a Riemannian metric and is not hyperbolic for a Lorentzian metric. Namely let W be a vector field on M with compact support. Let $(\Phi_s)_s$ be the flow of W, i.e. every $\Phi_s\colon M \to M$ is a diffeomorphism, $\Phi_0 = \mathrm{id}_M$ and $\frac{\partial \Phi_s(x)}{\partial s}|_{s=0} = W(x)$ for all $x \in M$. Let g be a semi-Riemannian metric on M and define $g_s := \Phi_s^* g$. Then $g_0 = g$. In order to compute $h := \frac{\partial}{\partial s}|_{s=0} g_s$ we take $X \in T_pM$, $p \in M$, we choose a curve c in M such that $c'(0) = X$ and we compute

$$\begin{aligned}
\frac{\partial}{\partial s}\Big|_{s=0} g_s(X,X) &= \frac{\partial}{\partial s}\Big|_{s=0} g(d\Phi_s(X), d\Phi_s(X)) \\
&= 2g\Big(\frac{\nabla}{\partial s}\Big|_{s=0} d\Phi_s(X), X\Big) \\
&= 2g\Big(\frac{\nabla}{\partial s}\Big|_{s=0} \frac{\partial}{\partial t}\Big|_{t=0} \Phi_s(c(t)), X\Big) \\
&= 2g\Big(\frac{\nabla}{\partial t}\Big|_{t=0} \frac{\partial}{\partial s}\Big|_{s=0} \Phi_s(c(t)), X\Big) \\
&= 2g\Big(\frac{\nabla}{\partial t}\Big|_{t=0} W(c(t)), X\Big) \\
&= 2g(\nabla_X W, X),
\end{aligned}$$

where we have interchanged the derivatives with respect to s and t since ∇ is torsion free. By polarization we get for all $X, Y \in T_pM$, $p \in M$

$$h(X,Y) = g(\nabla_X W, Y) + g(\nabla_Y W, X). \tag{6}$$

The Ricci curvature is a natural functional on the space of metrics, i.e. for every diffeomorphism $\Phi\colon M \to M$ we have $\mathrm{ric}_{\Phi^* g} = \Phi^* \mathrm{ric}_g$. Thus we get

$$\mathrm{ric}' = \frac{d}{ds}\Big|_{s=0} \mathrm{ric}_{g_s} = \frac{d}{ds}\Big|_{s=0} \Phi_s^* \mathrm{ric}.$$

If we repeat the above computation with ric instead of g we obtain

$$\mathrm{ric}'(X,Y) = \mathrm{ric}(\nabla_X W, Y) + \mathrm{ric}(\nabla_Y W, X).$$

Therefore if $\mathrm{ric} = 0$, then we have $\mathrm{ric}' = 0$ for every h of the form (6). In particular the kernel of the operator on the right hand side of Lemma 5.3 is infinite dimensional. Thus for a Riemannian metric this operator cannot be elliptic since otherwise for compact M its kernel would have finite dimension. We also see that for Lorentzian metrics this operator cannot be hyperbolic since otherwise its principal symbol would be given by the metric and then also in the Riemannian case its principal symbol would be given by the metric which is impossible.
Before we compute the first variation of the scalar curvature we need the following lemma.

Lemma 5.4. *Let (M, g) be a semi-Riemannian manifold, let h be a symmetric $(0,2)$-tensor field on M and let ω be a 1-form on M. Then we have*

(1) $\Delta \langle g, h\rangle = \langle g, \Delta_L h\rangle$,
(2) $\langle g, \operatorname{div}^* \omega\rangle = \delta\omega$.

Proof. (1) Let $p \in M$ and let $(E_i)_{i=1}^n$ be a local orthonormal frame such that we have $\nabla E_j|_p = 0$ for all j. Then at p we have

$$\begin{aligned}
\Delta \langle g, h\rangle &= \sum_{i=1}^n \varepsilon_i\{-\partial_{E_i}\partial_{E_i} \langle g, h\rangle + \partial_{\nabla_{E_i}E_i} \langle g, h\rangle\} \\
&= -\sum_{i,j=1}^n \varepsilon_i\varepsilon_j \partial_{E_i}\partial_{E_i} h(E_j, E_j) \\
&= -\sum_{i,j=1}^n \varepsilon_i\varepsilon_j\{(\nabla_{E_i}\nabla_{E_i} h)(E_j, E_j) + h(\nabla_{E_i}\nabla_{E_i}E_j, E_j) + h(E_j, \nabla_{E_i}\nabla_{E_i}E_j)\} \\
&= -\sum_{i,j=1}^n \varepsilon_i\varepsilon_j(\nabla_{E_i}\nabla_{E_i} h)(E_j, E_j) \\
&\quad -\sum_{i,j,k=1}^n \varepsilon_i\varepsilon_j\varepsilon_k\{\left\langle \nabla_{E_i}\nabla_{E_i}E_j, E_k\right\rangle + \left\langle E_k, \nabla_{E_i}\nabla_{E_i}E_j\right\rangle\} h(E_j, E_k) \\
&= -\sum_{i,j=1}^n \varepsilon_i\varepsilon_j(\nabla_{E_i}\nabla_{E_i} h)(E_j, E_j) - \sum_{i,j,k=1}^n \varepsilon_i\varepsilon_j\varepsilon_k\{\partial_{E_i}\partial_{E_i}\left\langle E_j, E_k\right\rangle\} h(E_j, E_k) \\
&= -\sum_{i,j=1}^n \varepsilon_i\varepsilon_j(\nabla_{E_i}\nabla_{E_i} h)(E_j, E_j) \\
&= \langle g, \nabla^*\nabla h\rangle .
\end{aligned}$$

By definition we have $\Delta_L h = \nabla^*\nabla h + \operatorname{ric}\circ h + h \circ \operatorname{ric} - 2\mathring{R}h$ and thus

$$\begin{aligned}
\langle g, \Delta_L h\rangle &= \langle g, \nabla^*\nabla h\rangle + \sum_{i,j=1}^n \varepsilon_i\varepsilon_j\{2\operatorname{ric}(E_j, E_i)h(E_j, E_i) - 2h(R(E_i, E_j)E_j, E_i)\} \\
&= \langle g, \nabla^*\nabla h\rangle + 2\langle \operatorname{ric}, h\rangle - 2\sum_{i,j,k=1}^n \varepsilon_i\varepsilon_j\varepsilon_k \left\langle R(E_i, E_j)E_j, E_k\right\rangle h(E_k, E_i) \\
&= \langle g, \nabla^*\nabla h\rangle .
\end{aligned}$$

(2) We have

$$\langle g, \operatorname{div}^* \omega\rangle = -\tfrac{1}{2}\sum_{i=1}^n \varepsilon_i\{(\nabla_{E_i}\omega)(E_i) + (\nabla_{E_i}\omega)(E_i)\} = -\sum_{i=1}^n \varepsilon_i(\nabla_{E_i}\omega)(E_i) = \delta\omega.$$

This finishes the proof. □

We compute the first variation of the scalar curvature.

Lemma 5.5. *We have*

$$\mathrm{scal}' = \Delta \langle g, h\rangle - \delta \mathrm{div}\, h - \langle \mathrm{ric}, h\rangle \,,$$

where δ is the codifferential acting on 1*-forms.*

Proof. By definition we have for all s

$$\mathrm{scal}^s = \sum_{i,j=1}^{n} (g^s)^{ij} \mathrm{ric}^s_{ij}$$

and thus

$$\mathrm{scal}' = -\sum_{i,j,k,\ell} g^{ik} g^{j\ell} h_{k\ell} \mathrm{ric}_{ij} + \sum_{i,j=1}^{n} g^{ij} \mathrm{ric}'_{ij} = -\langle \mathrm{ric}, h\rangle + \langle g, \mathrm{ric}'\rangle \,.$$

We note that

$$\langle g, \nabla d \langle g, h\rangle\rangle = \sum_{i=1}^{n} \varepsilon_i (\nabla_{E_i} d \langle g, h\rangle)(E_i) = \sum_{i=1}^{n} \varepsilon_i \partial_{E_i} \partial_{E_i} \langle g, h\rangle = -\Delta \langle g, h\rangle \,.$$

Therefore by Lemma 5.3 and Lemma 5.4 we get

$$\begin{aligned}
\langle g, \mathrm{ric}'\rangle &= \tfrac{1}{2} \langle g, \Delta_L h\rangle - \langle g, \mathrm{div}^* \mathrm{div}\, h\rangle - \tfrac{1}{2} \langle g, \nabla d \langle g, h\rangle\rangle \\
&= \tfrac{1}{2} \Delta \langle g, h\rangle - \delta \mathrm{div}\, h + \tfrac{1}{2} \Delta \langle g, h\rangle \\
&= \Delta \langle g, h\rangle - \delta \mathrm{div}\, h.
\end{aligned}$$

This finishes the proof. □

5.2. Constraint equations

Let M be a time-oriented Lorentzian manifold, let $\widehat{M} \subset M$ be a spacelike hypersurface and let ν be the future-directed unit normal vector field along $\widehat{M}$. We denote the metric on M by g and its restriction to $\widehat{M}$ by $\widehat{g}$. Furthermore we denote the Levi-Civita connections on M and $\widehat{M}$ by ∇ and $\widehat{\nabla}$, respectively. For vector fields X, Y on $\widehat{M}$ we have the decomposition

$$\nabla_X Y = \widehat{\nabla}_X Y + \widehat{K}(X, Y)\nu.$$

Here $\widehat{K}$ is a symmetric $(0, 2)$-tensor field on $\widehat{M}$. More precisely $\widehat{K}$ is the *second fundamental form* of $\widehat{M}$ in M. Namely taking the scalar product with ν in the above equation and using that ν is timelike we get $\langle \nabla_X Y, \nu\rangle = -\widehat{K}(X, Y)$. On the other hand we have

$$\langle \nabla_X Y, \nu\rangle = \partial_X \langle Y, \nu\rangle - \langle Y, \nabla_X \nu\rangle = -\langle Y, \nabla_X \nu\rangle$$

and thus $\widehat{K}(X,Y) = \langle \nabla_X \nu, Y\rangle$. From differential geometry we know that the curvature tensors of M and $\widehat{M}$ are related by the Gauss equation and the Codazzi equation. Namely let $X, Y, Z, W \in T_p\widehat{M}$, $p \in \widehat{M}$. Then the Gauss equation reads

$$\langle R(X,Y)Z, W\rangle = \left\langle \widehat{R}(X,Y)Z, W\right\rangle + \widehat{K}(X,W)\widehat{K}(Y,Z) - \widehat{K}(X,Z)\widehat{K}(Y,W)$$

and the Codazzi equation reads

$$\langle R(X,Y)\nu, Z\rangle = (\widehat{\nabla}_X \widehat{K})(Y,Z) - (\widehat{\nabla}_Y \widehat{K})(X,Z).$$

Let $E_1, ..., E_n$ be a local orthonormal frame on $\widehat{M}$ which is synchronous at $p \in \widehat{M}$, i.e. $\nabla E_i|_p = 0$. By contracting the Codazzi equation we get

$$\begin{aligned}\operatorname{ric}(X,\nu) &= \sum_{i=1}^{n-1} \langle R(E_i, X)\nu, E_i\rangle - \langle R(\nu, X)\nu, \nu\rangle = \sum_{i=1}^{n-1} \{(\widehat{\nabla}_{E_i}\widehat{K})(X, E_i) - (\widehat{\nabla}_X \widehat{K})(E_i, E_i)\} \\ &= \operatorname{div} \widehat{K}(X) - \partial_X \left\langle \widehat{g}, \widehat{K}\right\rangle\end{aligned}$$

and therefore $\operatorname{ric}(\cdot, \nu)|_{T\widehat{M}} = \operatorname{div} \widehat{K} - d\left\langle \widehat{g}, \widehat{K}\right\rangle$. Next we contract the Gauss equation twice. The terms on the right hand side are as follows

$$\begin{aligned}\sum_{i,j=1}^{n-1} \left\langle \widehat{R}(E_i, E_j)E_j, E_i\right\rangle &= \widehat{\operatorname{scal}}, \\ \sum_{i,j=1}^{n-1} \widehat{K}(E_i, E_i)\widehat{K}(E_j, E_j) &= \left\langle \widehat{g}, \widehat{K}\right\rangle^2, \\ \sum_{i,j=1}^{n-1} \widehat{K}(E_i, E_j)\widehat{K}(E_j, E_i) &= |\widehat{K}|^2\end{aligned}$$

and on the left hand side we get

$$\sum_{i,j=1}^{n-1} \left\langle R(E_i, E_j)E_j, E_i\right\rangle = \operatorname{scal} + \sum_{i=1}^{n-1} \langle R(E_i, \nu)\nu, E_i\rangle + \sum_{j=1}^{n-1} \left\langle R(\nu, E_j)E_j, \nu\right\rangle = \operatorname{scal} + 2\operatorname{ric}(\nu,\nu).$$

Thus from the Gauss equation we get

$$\operatorname{scal} + 2\operatorname{ric}(\nu,\nu) = \widehat{\operatorname{scal}} + \left\langle \widehat{g}, \widehat{K}\right\rangle^2 - |\widehat{K}|^2.$$

Assume that M is a vacuum solution, i.e. $\operatorname{ric} \equiv 0$. Then we get the so-called *constraint equations*

$$\widehat{\operatorname{scal}} + \left\langle \widehat{g}, \widehat{K}\right\rangle^2 - |\widehat{K}|^2 = 0 \tag{7}$$

$$\operatorname{div} \widehat{K} - d\left\langle \widehat{g}, \widehat{K}\right\rangle = 0. \tag{8}$$

The first equation is also called the *Hamiltonian constraint* and the second one the *momentum constraint*.

We note that the metric $\widehat{g}$ and the tensor field $\widehat{K}$ on $\widehat{M}$ can be regarded as initial data for the equation ric $\equiv 0$, where $\widehat{K}$ plays the role of the derivative with respect to time. Namely for $t \in \mathbb{R}$ such that $|t|$ is small we consider the map

$$\Phi_t : \quad \widehat{M} \to M, \quad p \mapsto \exp_p(tv),$$

and we define $\widehat{M}_t := \{\exp_p(tv) \,|\, p \in \widehat{M}\}$. We assume that there exists $t_0 > 0$ such that for all t with $|t| < t_0$ the map Φ_t is a diffeomorphism $\widehat{M} \to \widehat{M}_t$. This hypothesis is not very restrictive since it is satisfied for example if we replace $\widehat{M}$ by a compact subset of $\widehat{M}$. For $|t| < t_0$ we define $\widehat{g}_t := \Phi_t^* g$. Then we have $\widehat{g}_0 = g$ and the family $(\widehat{g}_t)_t$ of metrics determines the metric g in an open neighborhood of $\widehat{M}$ in M by

$$g = -dt^2 + \widehat{g}_t.$$

Let $X \in T_p\widehat{M}$, $p \in \widehat{M}$, and let c be a curve such that $c'(0) = X$. We compute

$$\begin{aligned}
\frac{\partial}{\partial t}\Big|_{t=0} \widehat{g}_t(X,X) &= \frac{\partial}{\partial t}\Big|_{t=0} g(d\Phi_t(X), d\Phi_t(X)) \\
&= 2\widehat{g}\Big(\frac{\nabla}{\partial t}\Big|_{t=0} d\Phi_t(X), X\Big) \\
&= 2\widehat{g}\Big(\frac{\nabla}{\partial t}\Big|_{t=0} \frac{\partial}{\partial s}\Big|_{s=0} \exp_{c(s)}(tv), X\Big) \\
&= 2\widehat{g}\Big(\frac{\nabla}{\partial s}\Big|_{s=0} \frac{\partial}{\partial t}\Big|_{t=0} \exp_{c(s)}(tv), X\Big) \\
&= 2\widehat{g}\Big(\frac{\nabla}{\partial s}\Big|_{s=0} v(c(s)), X\Big) \\
&= 2\widehat{g}(\nabla_X v, X) \\
&= 2\widehat{K}(X,X),
\end{aligned}$$

where we have interchanged the derivatives with respect to s and t since ∇ is torsion free. We conclude that $\frac{\partial}{\partial t}|_{t=0}\widehat{g}_t = 2\widehat{K}$. Thus $\widehat{g}$, $\widehat{K}$ can be regarded as initial data for the equation ric $\equiv 0$, where $\widehat{K}$ is the derivative with respect to time. If this initial value problem has a solution then $\widehat{g}$ and $\widehat{K}$ must satisfy the constraint equations (7), (8). In particular, the value of the unknown at $t = 0$ and its derivative with respect to time at $t = 0$ cannot be prescribed independently of each other. On the other hand, it follows from work by Y. Choquet-Bruhat that for all data $\widehat{g}$, $\widehat{K}$ on a Riemannian manifold $\widehat{M}$ satisfying the constraint equations the initial value problem has a solution, at least locally. The construction of solutions to the constraint equations (7), (8) is an active area of research.

5.3. Linearization of the constraint equations

We compute the linearization of the constraint equations (7), (8) at a given solution $(\widehat{g}, \widehat{K})$. We write

$$\widehat{g}_s = \widehat{g} + s\widehat{h} + O(s^2), \quad \widehat{K}_s = \widehat{K} + s\widehat{k} + O(s^2).$$

First we compute the linearization of the Hamiltonian constraint. By Lemma 5.5 we get

$$\widehat{\mathrm{scal}}' = \Delta\left\langle \widehat{g}, \widehat{h}\right\rangle - \delta\mathrm{div}\,\widehat{h} - \left\langle \widehat{\mathrm{ric}}, \widehat{h}\right\rangle.$$

We compute

$$\frac{\partial}{\partial s}\Big|_{s=0}\left\langle \widehat{g}_s, \widehat{K}_s\right\rangle = \frac{\partial}{\partial_s}\Big|_{s=0}\widehat{g}_s^{\alpha\beta}\widehat{K}_{s,\alpha\beta} = \widehat{g}^{\alpha\beta}\widehat{k}_{\alpha\beta} - \widehat{g}^{\alpha\mu}\widehat{g}^{\beta\nu}h_{\mu\nu}\widehat{K}_{\alpha\beta} = \left\langle \widehat{g}, \widehat{k}\right\rangle - \left\langle \widehat{K}, \widehat{h}\right\rangle \tag{9}$$

and thus we get

$$\frac{\partial}{\partial s}\Big|_{s=0}\left\langle \widehat{g}_s, \widehat{K}_s\right\rangle^2 = 2\left\langle \widehat{g}, \widehat{K}\right\rangle\left(\left\langle \widehat{g}, \widehat{k}\right\rangle - \left\langle \widehat{K}, \widehat{h}\right\rangle\right).$$

We also compute

$$\begin{aligned}
\frac{\partial}{\partial s}\Big|_{s=0}|\widehat{K}_s|^2 &= \frac{\partial}{\partial s}\Big|_{s=0}\widehat{g}_s^{\alpha\gamma}\widehat{g}_s^{\beta\delta}\widehat{K}_{s,\alpha\beta}\widehat{K}_{s,\gamma\delta}\\
&= -2\widehat{g}^{\alpha\mu}\widehat{g}^{\gamma\nu}\widehat{h}_{\mu\nu}\widehat{g}^{\beta\delta}\widehat{K}_{\alpha\beta}\widehat{K}_{\gamma\delta} + 2\widehat{g}^{\alpha\gamma}\widehat{g}^{\beta\delta}\widehat{k}_{\alpha\beta}\widehat{K}_{\gamma\delta}\\
&= -2\widehat{g}^{\alpha\mu}\widehat{g}^{\gamma\nu}\widehat{h}_{\mu\nu}(\widehat{K}\circ\widehat{K})_{\alpha\gamma} + 2\left\langle \widehat{k}, \widehat{K}\right\rangle\\
&= -2\left\langle \widehat{h}, \widehat{K}\circ\widehat{K}\right\rangle + 2\left\langle \widehat{k}, \widehat{K}\right\rangle.
\end{aligned}$$

Altogether we obtain the linearization of the Hamiltonian constraint (7) at $(\widehat{g}, \widehat{K})$

$$\begin{aligned}
0 &= \Delta\left\langle \widehat{g}, \widehat{h}\right\rangle - \delta\mathrm{div}\,\widehat{h} - \left\langle \widehat{\mathrm{ric}}, \widehat{h}\right\rangle + 2\left\langle \widehat{g}, \widehat{K}\right\rangle\left(\left\langle \widehat{g}, \widehat{k}\right\rangle - \left\langle \widehat{K}, \widehat{h}\right\rangle\right) + 2\left\langle \widehat{h}, \widehat{K}\circ\widehat{K}\right\rangle - 2\left\langle \widehat{k}, \widehat{K}\right\rangle\\
&= \Delta\left\langle \widehat{g}, \widehat{h}\right\rangle - \delta\mathrm{div}\,\widehat{h} + \left\langle -\widehat{\mathrm{ric}} - 2\left\langle \widehat{g}, \widehat{K}\right\rangle\widehat{K} + 2\widehat{K}\circ\widehat{K}, \widehat{h}\right\rangle + 2\left\langle\left\langle \widehat{g}, \widehat{K}\right\rangle\widehat{g} - \widehat{K}, \widehat{k}\right\rangle.
\end{aligned} \tag{10}$$

We note that this equation contains second order derivatives of $\widehat{h}$ but no derivatives of $\widehat{k}$. Next we compute the linearization of the momentum constraint. We take the differential on both sides of (9) and we get

$$\frac{\partial}{\partial s}\Big|_{s=0}d\left\langle \widehat{g}_s, \widehat{K}_s\right\rangle = d\left\langle \widehat{g}, \widehat{k}\right\rangle - d\left\langle \widehat{K}, \widehat{h}\right\rangle.$$

Thus the linearization of the momentum constraint (8) is

$$0 = \mathrm{div}\,\widehat{k} + d\left\langle \widehat{K}, \widehat{h}\right\rangle - d\left\langle \widehat{g}, \widehat{k}\right\rangle + \frac{\partial}{\partial s}\Big|_{s=0}\mathrm{div}_{\,\widehat{g}_s}\widehat{K}.$$

It remains to compute the variation of the divergence. We choose a family of local frames $((E_i^s)_{i=1}^{n-1})_s$ depending smoothly on s such that for all s the frame $(E_i^s)_{i=1}^{n-1}$ is orthonormal with respect to $\widehat{g}_s$ and such that at a fixed point $p \in \widehat{M}$ we have $\nabla E_i^s|_p = 0$. We define $e_j := \frac{\partial}{\partial s}|_{s=0}E_j^s$, $j = 1, ..., n-1$. In the equation $\varepsilon_i\delta_{ij} = \widehat{g}_s(E_i^s, E_j^s)$ we take the derivative with respect to s at $s = 0$ and we get

$$0 = \widehat{h}(E_i, E_j) + \widehat{g}(e_i, E_j) + \widehat{g}(E_i, e_j). \tag{11}$$

Multiplying by $\varepsilon_j E_j$ and summing over j we have

$$0 = \sum_{j=1}^{n-1}\varepsilon_j\widehat{h}(E_i, E_j)E_j + e_i + \sum_{j=1}^{n-1}\varepsilon_j\widehat{g}(E_i, e_j)E_j. \tag{12}$$

In the equation $0 = \nabla^s_{E_i^s} E_j^s$ we take the derivative with respect to s at $s = 0$ and we get

$$0 = \nabla'(E_i, E_j) + \nabla_{E_i} e_j.$$

Taking the inner product with E_i and using the variation of ∇ from Lemma 5.1 we get

$$\begin{aligned}
\widehat{g}(E_i, \nabla_{E_i} e_j) &= -\widehat{g}(\nabla'(E_i, E_j), E_i) \\
&= -\tfrac{1}{2}\{(\nabla_{E_i}\widehat{h})(E_i, E_j) + (\nabla_{E_j}\widehat{h})(E_i, E_i) - (\nabla_{E_i}\widehat{h})(E_i, E_j)\} \\
&= -\tfrac{1}{2}(\nabla_{E_j}\widehat{h})(E_i, E_i).
\end{aligned} \tag{13}$$

Let $X \in T_p\widehat{M}$ and extend X to a locally defined vector field such that $\nabla X|_p = 0$. Using the formula (12) for e_i and the variation of ∇ from Lemma 5.1 we compute

$$\begin{aligned}
&\frac{\partial}{\partial s}\Big|_{s=0} (\operatorname{div}_{g_s} \widehat{K})(X) = \frac{\partial}{\partial s}\Big|_{s=0} \sum_{i=1}^{n-1} \varepsilon_i (\nabla^s_{E_i^s}\widehat{K})(E_i^s, X) \\
&= \frac{\partial}{\partial s}\Big|_{s=0} \sum_{i=1}^{n-1} \varepsilon_i \{\partial_{E_i}\widehat{K}(E_i^s, X) - \widehat{K}(E_i^s, \nabla^s_{E_i^s} X)\} \\
&= \sum_{i=1}^{n-1} \varepsilon_i \{\partial_{e_i}\widehat{K}(E_i, X) + \partial_{E_i}\widehat{K}(e_i, X) - \widehat{K}(E_i, \nabla'(E_i, X))\} \\
&= \sum_{i,j=1}^{n-1} \varepsilon_i\varepsilon_j \{-\widehat{h}(E_i, E_j)\partial_{E_j}K(E_i, X) - \widehat{g}(E_i, e_j)\partial_{E_j}K(E_i, X)\} \\
&+ \sum_{i,j=1}^{n-1} \varepsilon_i\varepsilon_j \{\partial_{E_i}\{-\widehat{h}(E_i, E_j)\widehat{K}(E_j, X) - \widehat{g}(E_i, e_j)\widehat{K}(E_j, X)\} - \widehat{g}(\nabla'(E_i, X), E_j)\widehat{K}(E_i, E_j)\} \\
&= \sum_{i,j=1}^{n-1} \varepsilon_i\varepsilon_j \{-\widehat{h}(E_i, E_j)(\nabla_{E_j}K)(E_i, X) - \widehat{g}(E_i, e_j)(\nabla_{E_j}K)(E_i, X)\} \\
&+ \sum_{i,j=1}^{n-1} \varepsilon_i\varepsilon_j \{-(\nabla_{E_i}\widehat{h})(E_i, E_j)\widehat{K}(E_j, X) - \widehat{h}(E_i, E_j)(\nabla_{E_i}\widehat{K})(E_j, X)\} \\
&+ \sum_{i,j=1}^{n-1} \varepsilon_i\varepsilon_j \{-\widehat{g}(E_i, \nabla_{E_i} e_j)\widehat{K}(E_j, X) - \widehat{g}(E_i, e_j)(\nabla_{E_i}\widehat{K})(E_j, X)\} \\
&- \tfrac{1}{2}\sum_{i,j=1}^{n-1} \varepsilon_i\varepsilon_j \{(\nabla_{E_i}\widehat{h})(X, E_j) + (\nabla_X\widehat{h})(E_i, E_j) - (\nabla_{E_j}\widehat{h})(E_i, X)\}\widehat{K}(E_i, E_j)
\end{aligned}$$

Renaming i and j we see that two terms in the last line cancel. We also rename i and j in the last terms of line 2 and 3 on the right hand side and we get

$$\frac{\partial}{\partial s}\Big|_{s=0} (\operatorname{div}_{g_s} \widehat{K})(X)$$

$$= -\sum_{i,j=1}^{n-1} \varepsilon_i \varepsilon_j \{2\widehat{h}(E_i, E_j)(\nabla_{E_j}\widehat{K})(E_i, X) + (\nabla_{E_j}\widehat{K})(E_i, X)\{\widehat{g}(E_i, e_j) + \widehat{g}(E_j, e_i)\}\}$$
$$- \sum_{i,j=1}^{n-1} \varepsilon_i \varepsilon_j \Big\{(\nabla_{E_i}\widehat{h})(E_i, E_j)\widehat{K}(E_j, X) + \widehat{g}(E_i, \nabla_{E_i} e_j)\widehat{K}(E_j, X) - \tfrac{1}{2}(\nabla_X \widehat{h})(E_i, E_j)\widehat{K}(E_i, E_j)\Big\}$$

By (11) we have $\widehat{g}(E_i, e_j) + \widehat{g}(E_j, e_i) = -\widehat{h}(E_i, E_j)$. Using this and (13) we get

$$\frac{\partial}{\partial s}\Big|_{s=0}(\mathrm{div}_{g_s}\widehat{K})(X) = -\sum_{i,j=1}^{n-1} \varepsilon_i \varepsilon_j \{\widehat{h}(E_i, E_j)(\nabla_{E_j}\widehat{K})(E_i, X) + (\nabla_{E_i}\widehat{h})(E_i, E_j)\widehat{K}(E_j, X)\}$$
$$+ \tfrac{1}{2}\sum_{i,j=1}^{n-1} \varepsilon_i \varepsilon_j \{(\nabla_{E_j}\widehat{h})(E_i, E_i)\widehat{K}(E_j, X) - (\nabla_X \widehat{h})(E_i, E_j)\widehat{K}(E_i, E_j)\}.$$

We obtain

$$\frac{\partial}{\partial s}\Big|_{s=0}(\mathrm{div}_{g_s}\widehat{K})(X) = -\left\langle \widehat{h}, \nabla . \widehat{K}(\cdot, X)\right\rangle - \left\langle \mathrm{div}\,\widehat{h}, \widehat{K}(\cdot, X)\right\rangle$$
$$+ \tfrac{1}{2}\Big(\left\langle d\left\langle \widehat{g}, \widehat{h}\right\rangle, \widehat{K}(\cdot, X)\right\rangle - \left\langle \nabla_X \widehat{h}, \widehat{K}\right\rangle\Big).$$

Thus the linearization of the momentum constraint at $(\widehat{g}, \widehat{K})$ gives for all $X \in T\widehat{M}$

$$0 = (\mathrm{div}\,\widehat{k})(X) + \partial_X \left\langle \widehat{K}, \widehat{h}\right\rangle - \partial_X \left\langle \widehat{g}, \widehat{k}\right\rangle - \left\langle \widehat{h}, \nabla . \widehat{K}(\cdot, X)\right\rangle - \left\langle \mathrm{div}\,\widehat{h}, \widehat{K}(\cdot, X)\right\rangle$$
$$+ \tfrac{1}{2}\Big(\left\langle d\left\langle \widehat{g}, \widehat{h}\right\rangle, \widehat{K}(\cdot, X)\right\rangle - \left\langle \nabla_X \widehat{h}, \widehat{K}\right\rangle\Big). \tag{14}$$

Definition 5.6. A solution $(\widehat{g}, \widehat{K})$ to the constraint equations is called *linearization stable* if for every solution $(\widehat{h}, \widehat{k})$ of the linearization (10), (14) of the constraint equations at $(\widehat{g}, \widehat{K})$ there exists a continuously differentiable 1-parameter family $(\widehat{g}_s, \widehat{K}_s)$ of smooth solutions to the constraint equations, such that $\widehat{g}_0 = \widehat{g}$, $\widehat{K}_0 = \widehat{K}$, $\frac{\partial}{\partial s}|_{s=0}\widehat{g}_s = \widehat{h}$, $\frac{\partial}{\partial s}|_{s=0}\widehat{K}_s = \widehat{k}$.

The following example shows that not every solution to the constraint equations is linearization stable.

Example 5.7. Let $\widehat{M} = T^{n-1}$, the $(n-1)$-dimensional torus with a flat metric $\widehat{g}$ and let $\widehat{K} = 0$. We see immediately that $(\widehat{g}, \widehat{K})$ is a solution to the constraint equations (7), (8). In fact, the solution of this initial value problem gives the manifold $M = \mathbb{R} \times \widehat{M}$ with the flat metric $g = -dt^2 + \widehat{g}$. The linearization (10), (14) of the constraint equations at $(\widehat{g}, \widehat{K})$ gives

$$\Delta\left\langle \widehat{g}, \widehat{h}\right\rangle - \delta \mathrm{div}\,\widehat{h} = 0,$$
$$\mathrm{div}\,\widehat{k} - d\left\langle \widehat{g}, \widehat{k}\right\rangle = 0.$$

Using Cartesian coordinates on the universal cover $\mathbb{R}^{n-1}$ of T^{n-1} we write $\widehat{g} = \sum_j dx^j \otimes dx^j$ and we put

$$\widehat{k} := dx^1 \otimes dx^2 + dx^2 \otimes dx^1, \quad \widehat{h} := 0.$$

Then we have $\nabla\widehat{k} = 0$ and thus $\operatorname{div}\widehat{k} = 0$. We also have $\left\langle \widehat{g}, \widehat{k} \right\rangle = 0$ and thus $(\widehat{h}, \widehat{k})$ satisfy the linearization of the constraint equations at $(\widehat{g}, \widehat{K})$. Assume that there is a smooth 1-parameter family $(\widehat{g}_s, \widehat{K}_s)$ of solutions to the constraint equations such that $\widehat{g}_0 = \widehat{g}$, $\widehat{K}_0 = \widehat{K}$, $\frac{\partial}{\partial s}|_{s=0}\widehat{g}_s = \widehat{h}$, $\frac{\partial}{\partial s}|_{s=0}\widehat{K}_s = \widehat{k}$. Then we have

$$\begin{aligned}
\left\langle \widehat{g}_s, \widehat{K}_s \right\rangle_{\widehat{g}_s} &= \left\langle \widehat{g} + O(s^2), s\widehat{k} + O(s^2) \right\rangle_{\widehat{g}+O(s^2)} = s\left\langle \widehat{g}, \widehat{k} \right\rangle + O(s^2) = O(s^2),\\
\left|\widehat{K}_s\right|^2_{\widehat{g}_s} &= \left|s\widehat{k} + O(s^2)\right|_{\widehat{g}_s+O(s^2)} = s^2\left|\widehat{k}\right|^2_{\widehat{g}} + O(s^3)
\end{aligned}$$

and thus from the Hamiltonian constraint (7) we get

$$\widehat{\mathrm{scal}}_{\widehat{g}_s} = -\left\langle \widehat{g}_s, \widehat{K}_s \right\rangle^2_{\widehat{g}_s} + \left|\widehat{K}_s\right|^2_{\widehat{g}_s} = O(s^4) + s^2\left|\widehat{k}\right|^2_{\widehat{g}} + O(s^3) = 2s^2 + O(s^3) = s^2(2 + O(s)).$$

For every s the term $O(s)$ is bounded uniformly in T^{n-1} since T^{n-1} is compact. Thus for small $|s|$, $s \neq 0$ we have $\widehat{\mathrm{scal}}_{\widehat{g}_s} > 0$ everywhere on T^{n-1}. This contradicts the fact that T^{n-1} does not admit a Riemannian metric of positive scalar curvature. This follows from the Gauss-Bonnet theorem for $n-1 = 2$ and was proved by Schoen-Yau [**?**, Cor. 2] for $n-1 \leq 7$ and by Gromov-Lawson [**?**, Cor. A, p. 94] for general n. Thus for $(\widehat{h}, \widehat{k})$ there is no smooth 1-parameter family of solutions to the constraint equations as in the above definition and therefore the solution $(\widehat{g}, \widehat{K})$ is not linearization stable.

Next we investigate under which assumption one can conclude that solutions to the constraint equations are linearization stable. Using the operators

$$\begin{aligned}
\mathcal{H}: &\quad \Gamma(\odot^2 T^*\widehat{M}) \times \Gamma(\odot^2 T^*\widehat{M}) \to C^\infty(\widehat{M}), &\quad \mathcal{H}(\widehat{g}, \widehat{K}) &:= \widehat{\mathrm{scal}} + \left\langle \widehat{g}, \widehat{K} \right\rangle^2 - |\widehat{K}|^2,\\
\mathcal{M}: &\quad \Gamma(\odot^2 T^*\widehat{M}) \times \Gamma(\odot^2 T^*\widehat{M}) \to \Omega^1(\widehat{M}), &\quad \mathcal{M}(\widehat{g}, \widehat{K}) &:= \operatorname{div}\widehat{K} - \left\langle \widehat{g}, \widehat{K} \right\rangle,
\end{aligned}$$

the constraint equations (7), (8) read $\mathcal{H}(\widehat{g}, \widehat{K}) = 0$, $\mathcal{M}(\widehat{g}, \widehat{K}) = 0$. By (10) and (14) the differentials of $\mathcal{H}$ and $\mathcal{M}$ take the form

$$\begin{aligned}
d_{(\widehat{g},\widehat{K})}\mathcal{H}(\widehat{h}, \widehat{k}) &= \Delta\left\langle \widehat{g}, \widehat{h} \right\rangle - \delta\operatorname{div}\widehat{h} + \text{terms of order 0 in } \widehat{k},\\
d_{(\widehat{g},\widehat{K})}\mathcal{M}(\widehat{h}, \widehat{k}) &= \operatorname{div}\widehat{k} - d\left\langle \widehat{g}, \widehat{k} \right\rangle + \text{terms of order at most 1 in } \widehat{h}.
\end{aligned}$$

Thus the operator

$$d\mathcal{H} := d_{(\widehat{g},\widehat{K})}\mathcal{H}: \quad \Gamma(\odot^2 T^*\widehat{M}) \oplus \Gamma(\odot^2 T^*\widehat{M}) \to C^\infty(\widehat{M})$$

is of order 2 in $\widehat{h}$ and of order 0 in $\widehat{k}$ and the operator

$$d\mathcal{M} := d_{(\widehat{g},\widehat{K})}\mathcal{M}: \quad \Gamma(\odot^2 T^*\widehat{M}) \oplus \Gamma(\odot^2 T^*\widehat{M}) \to \Omega^1(\widehat{M})$$

is of order 1 in $\widehat{k}$ and of order at most 1 in $\widehat{h}$. It follows that

$$d\mathcal{H} \oplus d\mathcal{M}: \quad \Gamma(\odot^2 T^*\widehat{M}) \oplus \Gamma(\odot^2 T^*\widehat{M}) \to C^\infty(\widehat{M}) \oplus \Omega^1(\widehat{M})$$

is a linear differential operator of bi-degree $(2,1)$.
We recall that for every linear differential operator $P\colon \Gamma(E) \to \Gamma(F)$ of order ℓ, where E and F are vector bundles over $\widehat{M}$, the principal symbol of P is defined as follows: Let $p \in \widehat{M}$, $\xi \in T_p^*\widehat{M}$, $s \in E_p$. We choose $f \in C^\infty(\widehat{M})$ such that $f(p) = 0$ and $df|_p = \xi$ and we choose an extension of s to a locally defined section of E in a neighborhood of p, which is also denoted by s. Then we define

$$\sigma_\ell(P,\xi)(s) := P\Big(\frac{1}{\ell!} f^\ell s\Big)\Big|_p$$

and we note that $\sigma_\ell(P,\xi)(s)$ is independent of the choices of f and s. We note that the definitions in the literature vary slightly. In an analogous way we can define the principal symbol of a differential operator of bi-degree (k,ℓ). By this definition we have

$$\sigma_{2,1}(d\mathcal{H} \oplus d\mathcal{M}, \xi)(\widehat{h},\widehat{k}) = \begin{pmatrix} d\mathcal{H}(\frac{1}{2}f^2\widehat{h}, f\widehat{k}) \\ d\mathcal{M}(\frac{1}{2}f^2\widehat{h}, f\widehat{k}) \end{pmatrix}.$$

Using that $f(p) = 0$ and $df|_p = \xi$ we compute at p

$$\begin{aligned}
\Delta\left\langle \widehat{g}, \tfrac{1}{2}f^2\widehat{h}\right\rangle &= \tfrac{1}{2}\Delta\Big(f^2\left\langle \widehat{g},\widehat{h}\right\rangle\Big) \\
&= \tfrac{1}{2}\Delta(f^2)\left\langle \widehat{g},\widehat{h}\right\rangle - \tfrac{1}{2}\left\langle \nabla(f^2), \nabla\Big(\left\langle \widehat{g},\widehat{k}\right\rangle\Big)\right\rangle + \tfrac{1}{2}f^2\Delta\left\langle \widehat{g},\widehat{h}\right\rangle \\
&= (-\langle \nabla f, \nabla f\rangle + f\Delta f)\left\langle \widehat{g},\widehat{h}\right\rangle \\
&= -\left\langle \xi^\#, \xi^\#\right\rangle \left\langle \widehat{g},\widehat{h}\right\rangle \\
&= -|\xi|^2\left\langle \widehat{g},\widehat{h}\right\rangle,
\end{aligned}$$

$$\sigma_1(\operatorname{div},\xi)(\widehat{h}) = \operatorname{div}(f\widehat{h}) = \sum_{i=1}^{n-1} \nabla_{E_i}(f\widehat{h})(E_i,\cdot) = \sum_{i=1}^{n-1}(\partial_{E_i} f \cdot \widehat{h} + f\nabla_{E_i}\widehat{h})(E_i,\cdot) = \widehat{h}(\xi^\#,\cdot),$$

$$\sigma_1(\delta,\xi)(\omega) = \delta(f\omega) = f\delta\omega - \langle df,\omega\rangle = -\langle df,\omega\rangle = -\omega(\xi^\#),$$

$$d\left\langle \widehat{g}, f\widehat{k}\right\rangle = d\Big(f\left\langle \widehat{g},\widehat{k}\right\rangle\Big) = \left\langle \widehat{g},\widehat{k}\right\rangle df + f d\left\langle \widehat{g},\widehat{k}\right\rangle = \left\langle \widehat{g},\widehat{k}\right\rangle \xi.$$

Since the principal symbol is compatible with composition of differential operators we get

$$\sigma_2(\delta\operatorname{div},\xi)(\widehat{h}) = \sigma_1(\delta,\xi)\circ\sigma_1(\operatorname{div},\xi)(\widehat{h}) = -\widehat{h}(\xi^\#,\xi^\#).$$

Using these computations we obtain

$$\sigma_{2,1}(d\mathcal{H} \oplus d\mathcal{M}, \xi)(\widehat{h},\widehat{k}) = \begin{pmatrix} -|\xi|^2\left\langle \widehat{g},\widehat{h}\right\rangle + \widehat{h}(\xi^\#,\xi^\#) \\ \widehat{k}(\xi^\#,\cdot) - \left\langle \widehat{g},\widehat{k}\right\rangle \xi \end{pmatrix}.$$

Lemma 5.8. *For every $\xi \in T_p^*\widehat{M} \setminus \{0\}$ the operator*

$$\sigma_{2,1}(d\mathcal{H} \oplus d\mathcal{M}, \xi) : \quad \odot^2 T_p^*\widehat{M} \oplus \odot^2 T_p^*\widehat{M} \to \mathbb{R} \oplus T_p^*\widehat{M}$$

is surjective.

Proof. (a) Let $r \in \mathbb{R}$. We put $E_1 := \frac{\xi^\#}{|\xi|}$ and extend to an orthonormal basis $E_1, ..., E_{n-1}$ of $T_p\widehat{M}$ where $n-1 \geq 2$. Then we define $\widehat{h}$ such that $\widehat{h}(E_1, E_1) = \frac{r}{|\xi|^2}$, $\widehat{h}(E_2, E_2) = -\frac{r}{|\xi|^2}$ and $\widehat{h}(E_i, E_j) = 0$ otherwise. In particular, we have $\widehat{h}(\xi^\#, \xi^\#) = r$ and $\left\langle \widehat{g}, \widehat{h} \right\rangle = 0$. We also put $\widehat{k} = 0$. It follows that

$$\sigma_{2,1}(d\mathcal{H} \oplus d\mathcal{M}, \xi)(\widehat{h}, 0) = \begin{pmatrix} r \\ 0 \end{pmatrix}.$$

(b) Let $\omega \in T_p^*\widehat{M}$. *Case 1:* ω and ξ are linearly independent.
We put

$$\widehat{k} := \frac{|\omega|^2}{|\xi|^2|\omega|^2 - \langle \xi, \omega \rangle^2}(\omega \otimes \xi + \xi \otimes \omega) - \frac{\langle \omega, \xi \rangle}{|\xi|^2|\omega|^2 - \langle \xi, \omega \rangle^2}\omega \otimes \omega.$$

We also put $E_1 := \frac{\xi^\#}{|\xi|}$ and extend to an orthonormal basis $E_1, ..., E_{n-1}$ of $T_p\widehat{M}$. In particular, we have $\xi(E_j) = 0$ for all $j \geq 2$ and we get

$$\begin{aligned} \left\langle \widehat{g}, \widehat{k} \right\rangle &= \frac{|\omega|^2}{|\xi|^2|\omega|^2 - \langle \xi, \omega \rangle^2} \cdot 2\omega(E_1)\xi(E_1) - \frac{\langle \omega, \xi \rangle}{|\xi|^2|\omega|^2 - \langle \xi, \omega \rangle^2}|\omega|^2 \\ &= \frac{|\omega|^2}{|\xi|^2|\omega|^2 - \langle \xi, \omega \rangle^2} \cdot 2\langle \omega, \xi \rangle - \frac{\langle \omega, \xi \rangle}{|\xi|^2|\omega|^2 - \langle \xi, \omega \rangle^2}|\omega|^2 \\ &= \frac{\langle \omega, \xi \rangle}{|\xi|^2|\omega|^2 - \langle \xi, \omega \rangle^2}|\omega|^2 \end{aligned}$$

and since $\omega(\xi^\#) = \langle \omega, \xi \rangle$ and $\xi(\xi^\#) = |\xi|^2$ we obtain

$$\begin{aligned} \widehat{k}(\xi^\#, \cdot) - \left\langle \widehat{g}, \widehat{k} \right\rangle \xi &= \frac{|\omega|^2}{|\xi|^2|\omega|^2 - \langle \xi, \omega \rangle^2}(\langle \omega, \xi \rangle \xi + |\xi|^2\omega) - \frac{\langle \omega, \xi \rangle}{|\xi|^2|\omega|^2 - \langle \xi, \omega \rangle^2}\langle \omega, \xi \rangle \omega \\ &\quad - \frac{\langle \omega, \xi \rangle}{|\xi|^2|\omega|^2 - \langle \xi, \omega \rangle^2}|\omega|^2\xi \\ &= \frac{|\xi|^2|\omega|^2 - \langle \xi, \omega \rangle^2}{|\xi|^2|\omega|^2 - \langle \xi, \omega \rangle^2}\omega \\ &= \omega. \end{aligned}$$

Case 2: ω and ξ are linearly dependent.
We may assume without loss of generality that $\omega = \xi$. We put $E_1 := \frac{\xi^\#}{|\xi|}$ and extend to an orthonormal basis $E_1, ..., E_{n-1}$ of $T_p\widehat{M}$ where $n-1 \geq 2$. Then we define $\widehat{k}$ such that $\widehat{k}(E_1, E_1) = 1$,

$\widehat{k}(E_2, E_2) = -1$ and $\widehat{k}(E_i, E_j) = 0$ otherwise. In particular we have $\widehat{k}(\xi^\#, \cdot) = \xi = \omega$ and $\left\langle \widehat{g}, \widehat{k} \right\rangle = 0$.
In both cases we put $\widehat{h} = 0$. It follows that

$$\sigma_{2,1}(d\mathcal{H} \oplus d\mathcal{M}, \xi)(0, \widehat{k}) = \begin{pmatrix} 0 \\ \omega \end{pmatrix}$$

and this shows the assertion. □

Let V, W be finite dimensional Euclidean vector spaces and let $A\colon\ V \to W$ be linear and surjective. Then $AA^*\colon\ W \to W$ is bijective. Namely, if $y \in \ker(AA^*)$ then we have $0 = \langle AA^*y, y\rangle = |A^*y|^2$ and thus $A^*y = 0$. Since A is surjective we may write $y = Ax$ and we get $0 = \langle A^*Ax, x\rangle = |Ax|^2 = |y|^2$ and thus $y = 0$. It follows that AA^* is injective and thus bijective.
We conclude that the operator

$$\sigma_{2,1}(d\mathcal{H} \oplus d\mathcal{M}, \xi) \circ \sigma_{2,1}(d\mathcal{H} \oplus d\mathcal{M}, \xi)^* : \quad \mathbb{R} \oplus T_p^*\widehat{M} \to \mathbb{R} \oplus T_p^*\widehat{M}$$

is bijective. Since $d\mathcal{M}$ is an operator of degree 1 in $\widehat{k}$ we have

$$\sigma_{2,1}(d\mathcal{H} \oplus d\mathcal{M}, \xi)^* = \sigma_{2,1}((d\mathcal{H} \oplus (-d\mathcal{M}), \xi)^*)$$

and therefore

$$\begin{aligned} &\sigma_{2,1}(d\mathcal{H} \oplus d\mathcal{M}, \xi) \circ \sigma_{2,1}(d\mathcal{H} \oplus d\mathcal{M}, \xi)^* \\ &= \sigma_{4,2}((d\mathcal{H} \oplus d\mathcal{M}) \circ (d\mathcal{H} \oplus d\mathcal{M})^*, \xi) \circ \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \\ &= \sigma_{4,2}(d(\mathcal{H} \times \mathcal{M}) \circ d(\mathcal{H} \times \mathcal{M})^*, \xi) \circ \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \end{aligned}$$

Since this operator is bijective it follows that

$$P := d(\mathcal{H} \times \mathcal{M}) \circ d(\mathcal{H} \times \mathcal{M})^* : \quad C^\infty(\widehat{M}) \oplus \Omega^1(\widehat{M}) \to C^\infty(\widehat{M}) \oplus \Omega^1(\widehat{M})$$

is a linear elliptic differential operator of bi-degree $(4, 2)$. Obviously P is formally self-adjoint. In particular, if $\widehat{M}$ is compact, then we have $\dim\ker(P) = \operatorname{codim}\operatorname{im}(P) < \infty$. Furthermore all distributional solutions of the equation $P(f, \omega) = 0$ are smooth.
We note that we have $\ker(P) = \ker(d(\mathcal{H} \times \mathcal{M})^*)$. Indeed if $P(f, \omega) = 0$, then by taking the L^2-product with (f, ω) we get

$$0 = \int_{\widehat{M}} \langle d(\mathcal{H} \times \mathcal{M})d(\mathcal{H} \times \mathcal{M})^*(f, \omega), (f, \omega)\rangle\, d\widehat{\mathrm{vol}} = \int_{\widehat{M}} |d(\mathcal{H} \times \mathcal{M})^*(f, \omega)|^2 d\widehat{\mathrm{vol}}$$

and thus $d(\mathcal{H} \times \mathcal{M})^*(f, \omega) = 0$.
Now the principal symbol of the operator $d(\mathcal{H} \times \mathcal{M})^*$ is injective but not surjective. Such an operator is also called overdetermined elliptic.

Remark 5.9. One can compute that for a Lorentzian manifold M with $\mathrm{ric} \equiv 0$ and for a spacelike hypersurface $\widehat{M} \subset M$ with future pointing unit normal ν the following holds.

(1) One has

$$\ker(d(\mathcal{H} \times \mathcal{M})^*) = \left\{(f, \omega) \in C^\infty(\widehat{M}) \times \Omega^1(\widehat{M}) \,\middle|\, \begin{array}{l} \text{there is a Killing vector field } X \\ \text{on } M \text{ such that } X|_{\widehat{M}} = f\nu + \omega^\# \end{array}\right\}.$$

The elements of $\ker(d(\mathcal{H} \times \mathcal{M})^*)$ are sometimes called *Killing initial data* or *KIDs*.

(2) If $\widehat{M}$ is compact and $j > \frac{n}{2}$ where $n - 1 = \dim \widehat{M}$, then the map

$$\mathcal{H} \times \mathcal{M} : \quad W^{j+2,p}(\odot^2 T^*\widehat{M}) \times W^{j+2,p}(\odot^2 T^*\widehat{M}) \to W^{j,p}(\widehat{M}) \times W^{j+1,p}(T^*\widehat{M})$$

is smooth. Here $W^{j,p}(E)$ is the space of sections of a vector bundle E over $\widehat{M}$ whose weak derivatives up to order j exist and are in $L^p(\widehat{M})$.

Theorem 5.10 (Moncrief). *Let M be a time-oriented ricci-flat Lorentzian manifold without Killing vector fields. Let $\widehat{M} \subset M$ be a compact spacelike hypersurface. Then every solution of the constraint equations (7), (8) is linearization stable.*

Example 5.7 shows that the hypothesis that M has no Killing vector fields cannot be removed. Indeed, on $M = \mathbb{R} \times T^{n-1}$ with a flat Lorentzian metric parallel vector fields are Killing vector fields.

Proof. Since M has no Killing vector fields we know that $\ker(d(\mathcal{H} \times \mathcal{M})^*) = \{0\}$ and thus $d(\mathcal{H} \times \mathcal{M})$ is surjective. It follows that for all $j > \frac{n}{2}$ the map

$$\mathcal{H} \times \mathcal{M} : \quad W^{j+2,p}(\odot^2 T^*\widehat{M}) \times W^{j+2,p}(\odot^2 T^*\widehat{M}) \to W^{j,p}(\widehat{M}) \times W^{j+1,p}(T^*\widehat{M})$$

is a submersion. Since $\mathrm{im}(d(\mathcal{H}\times\mathcal{M})^*)$ is a complement to $\ker(d(\mathcal{H}\times\mathcal{M}))$ we may use the implicit function theorem for Banach spaces which tells us that $(\mathcal{H} \times \mathcal{M})^{-1}(0)$ is a smooth submanifold of $W^{j+2,p}(\widehat{M}) \times W^{j+2,p}(\widehat{M})$. Let $(\widehat{g}, \widehat{K})$ be a solution to the constraint equations (7), (8) and let $(\widehat{h}, \widehat{k})$ be a solution to the linearization (10), (14) of the constraint equations at $(\widehat{g}, \widehat{K})$. Then for all $j > \frac{n}{2}$ there is a C^1-curve

$$[-1, 1] \ni s \mapsto (\widehat{g}_s^{(j)}, \widehat{K}_s^{(j)}) \in (\mathcal{H} \times \mathcal{M})^{-1}(0) \subset W^{j+2,p}(\widehat{M}) \times W^{j+2,p}(\widehat{M})$$

such that $\widehat{g}_0^{(j)} = \widehat{g}$, $\widehat{K}_0^{(j)} = \widehat{K}$, $\frac{\partial}{\partial s}|_{s=0}\widehat{g}_s^{(j)} = \widehat{h}$, $\frac{\partial}{\partial s}|_{s=0}\widehat{K}_s^{(j)} = \widehat{k}$. After re-parametrization of the curves away from $s = 0$ we may assume that for all $j > \frac{n}{2}$ and for all s we have

$$\begin{aligned}
\left\|\widehat{g}_s^{(j)}\right\|_{W^{j+2,p}} &\le \left\|\widehat{g}\right\|_{W^{j+2,p}} + 1, \\
\left\|\widehat{K}_s^{(j)}\right\|_{W^{j+2,p}} &\le \left\|\widehat{K}\right\|_{W^{j+2,p}} + 1, \\
\left\|\partial_s\widehat{g}_s^{(j)}\right\|_{W^{j+2,p}} &\le \left\|\widehat{h}\right\|_{W^{j+2,p}} + 1, \\
\left\|\partial_s\widehat{K}_s^{(j)}\right\|_{W^{j+2,p}} &\le \left\|\widehat{k}\right\|_{W^{j+2,p}} + 1.
\end{aligned}$$

Then for every $j_0 > \frac{n}{2}$ the sequence $(\widehat{g}_\bullet^{(j)}, \widehat{K}_\bullet^{(j)})_j$ is eventually contained and bounded in $C^1([-1,1], W^{j_0+2,p})$. By Rellich's embedding theorem and after taking a diagonal subsequence we obtain a subsequence converging to an element $(\widehat{g}_\bullet^{(\infty)}, \widehat{K}_\bullet^{(\infty)})$ in every space $C^1([-1,1], W^{j_0+2,p})$, $j_0 > \frac{n}{2}$. By the Sobolev embedding theorem we conclude that $(\widehat{g}_\bullet^{(\infty)}, \widehat{K}_\bullet^{(\infty)}) \in C^1([-1,1], C^\infty(\widehat{M}))$. This 1-parameter family of solutions to (7), (8) has the desired properties.□

5.4. Existence of solutions to the gravitational wave equation

We start by proving the following lemma.

Lemma 5.11. *Let (M,g) be a semi-Riemannian manifold. Then for all symmetric $(0,2)$-tensor fields h on M the following holds.*

(1) *For all vector fields X, Y on M we have*

$$\nabla_X\nabla_Y h - \nabla_Y\nabla_X h - \nabla_{[X,Y]}h = -h(R(X,Y)\cdot,\cdot) - h(\cdot, R(X,Y)\cdot).$$

(2) *If* $\mathrm{ric} \equiv 0$, *then* $\nabla^*\nabla \mathrm{div}\, h = \mathrm{div}\,(\Delta_L h)$.

Proof. (1) We may assume that X, Y, X_1, X_2 are locally defined vector fields such that at a fixed point $p \in M$ we have $\nabla X_1|_p = 0$, $\nabla X_2|_p = 0$, $\nabla X|_p = 0$, $\nabla Y|_p = 0$. Then at p we have

$$\begin{aligned}
(\nabla_X\nabla_Y h)(X_1,X_2) &= \partial_X(\nabla_Y h)(X_1,X_2) - (\nabla_Y h)(\nabla_X X_1, X_2) - (\nabla_Y h)(X_1, \nabla_X X_2)\\
&= \partial_X\partial_Y h(X_1,X_2) - \partial_X h(\nabla_Y X_1, X_2) - \partial_X h(X_1, \nabla_Y X_2)\\
&\quad - \partial_Y h(\nabla_X X_1, X_2) + h(\nabla_Y\nabla_X X_1, X_2)\\
&\quad - \partial_Y h(X_1, \nabla_X X_2) + h(X_1, \nabla_Y\nabla_X X_2),\\
(\nabla_Y\nabla_X h)(X_1,X_2) &= \partial_Y\partial_X h(X_1,X_2) - \partial_Y h(\nabla_X X_1, X_2) - \partial_Y h(X_1, \nabla_X X_2)\\
&\quad - \partial_X h(\nabla_Y X_1, X_2) + h(\nabla_X\nabla_Y X_1, X_2)\\
&\quad - \partial_X h(X_1, \nabla_Y X_2) + h(X_1, \nabla_X\nabla_Y X_2).
\end{aligned}$$

Taking the difference and using that $[X,Y]|_p = 0$ we obtain the result of part (1).
(2) Since $\mathrm{ric} \equiv 0$ we have $\Delta_L h = \nabla^*\nabla h - 2\mathring{R}h$ where by definition

$$(\mathring{R}h)(X,Y) = \sum_{i=1}^n \varepsilon_i h(R(E_i,X)Y, E_i)$$

for $X, Y \in T_pM$, $p \in M$. Let X be a locally defined vector field on M and let $(E_i)_{i=1}^n$ be a local orthonormal frame such that at a fixed point $p \in M$ we have $\nabla X|_p = 0$ and $\nabla E_i|_p = 0$ for all i.

The following computations are valid at p. By the first and the second Bianchi identity we get

$$
\begin{aligned}
(\operatorname{div} \mathring{R}h)(X) &= \sum_{j=1}^{n} \varepsilon_j (\nabla_{E_j} \mathring{R}h)(X, E_j) \\
&= \sum_{i,j=1}^{n} \varepsilon_i \varepsilon_j \{(\nabla_{E_j} h)(R(E_i, X)E_j, E_i) + h((\nabla_{E_j} R)(E_i, X)E_j, E_i)\} \\
&= \sum_{i,j,k=1}^{n} \varepsilon_i \varepsilon_j \varepsilon_k \{\left\langle R(E_i, X)E_j, E_k \right\rangle (\nabla_{E_j} h)(E_k, E_i) + \left\langle (\nabla_{E_j} R)(E_i, X)E_j, E_k \right\rangle h(E_k, E_i)\} \\
&= \sum_{i,j,k=1}^{n} \varepsilon_i \varepsilon_j \varepsilon_k \{- \left\langle R(X, E_j)E_i, E_k \right\rangle - \left\langle R(E_j, E_i)X, E_k \right\rangle\} (\nabla_{E_j} h)(E_k, E_i) \\
&+ \sum_{i,j,k=1}^{n} \varepsilon_i \varepsilon_j \varepsilon_k \{- \left\langle (\nabla_{E_i} R)(X, E_j)E_j, E_k \right\rangle - \left\langle (\nabla_X R)(E_j, E_i)E_j, E_k \right\rangle\} h(E_k, E_i).
\end{aligned}
$$

Since $\left\langle R(X, E_j)E_i, E_k \right\rangle$ is antisymmetric in i, k and $(\nabla_{E_j} h)(E_k, E_i)$ is symmetric in i, k the first sum on the right hand side vanishes. The third and fourth sum on the right hand side are proportional to derivatives of ric and therefore vanish. Thus we get

$$
(\operatorname{div} \mathring{R}h)(X) = \sum_{i,j=1}^{n} \varepsilon_i \varepsilon_j (\nabla_{E_j} h)(R(E_i, E_j)X, E_i). \tag{15}
$$

Next we compute

$$
\begin{aligned}
(\operatorname{div} \nabla^* \nabla h)(X) &= \sum_{i=1}^{n} \varepsilon_i \partial_{E_i} (\nabla^* \nabla h)(X, E_i) \\
&= \sum_{i,j=1}^{n} \varepsilon_i \varepsilon_j \{-(\nabla_{E_i} \nabla_{E_j} \nabla_{E_j} h)(X, E_i) + (\nabla_{E_i} \nabla_{\nabla_{E_j} E_j} h)(X, E_i)\}.
\end{aligned}
$$

By part (1) and since $\nabla E_j|_p = 0$ we have for all i, j

$$
\begin{aligned}
(\nabla_{E_i} \nabla_{\nabla_{E_j} E_j} h)(X, E_i) &= (\nabla_{\nabla_{E_j} E_j} \nabla_{E_i} h)(X, E_i) + (\nabla_{[E_i, \nabla_{E_j} E_j]} h)(X, E_i) \\
&- h(R(E_i, \nabla_{E_j} E_j)X, E_i) - h(X, R(E_i, \nabla_{E_j} E_j)E_i) \\
&= (\nabla_{\nabla_{E_i} \nabla_{E_j} E_j} h)(X, E_i)
\end{aligned}
$$

and thus

$$
(\operatorname{div} \nabla^* \nabla h)(X) = \sum_{i,j=1}^{n} \varepsilon_i \varepsilon_j \{-(\nabla_{E_i} \nabla_{E_j} \nabla_{E_j} h)(X, E_i) + (\nabla_{\nabla_{E_i} \nabla_{E_j} E_j} h)(X, E_i)\}. \tag{16}
$$

Using part (1) with $\nabla_{E_j} h$ instead of h and using that $[E_i, E_j]|_p = 0$ we get for all i, j

$$
\begin{aligned}
(\nabla_{E_i} \nabla_{E_j} \nabla_{E_j} h)(X, E_i) &= (\nabla_{E_j} \nabla_{E_i} \nabla_{E_j} h)(X, E_i) \\
&- (\nabla_{E_j} h)(R(E_i, E_j)X, E_i) - (\nabla_{E_j} h)(X, R(E_i, E_j)E_i).
\end{aligned} \tag{17}
$$

By applying ∂_{E_j} to the equation in part (1) we get

$$
\begin{aligned}
(\nabla_{E_j}\nabla_{E_i}\nabla_{E_j}h)(X,E_i) &= \partial_{E_j}\{(\nabla_{E_j}\nabla_{E_i}h)(X,E_i) + (\nabla_{[E_i,E_j]}h)(X,E_i) \\
&- h(R(E_i,E_j)X,E_i) - h(X,R(E_i,E_j)E_i)\} \\
&= (\nabla_{E_j}\nabla_{E_j}\nabla_{E_i}h)(X,E_i) + (\nabla_{E_j}\nabla_{[E_i,E_j]}h)(X,E_i) \\
&- (\nabla_{E_j}h)(R(E_i,E_j)X,E_i) - h((\nabla_{E_j}R)(E_i,E_j)X,E_i) \\
&- (\nabla_{E_j}h)(X,R(E_i,E_j)E_i) - h(X,(\nabla_{E_j}R)(E_i,E_j)E_i).
\end{aligned}
$$

Since $[E_i,E_j]|_p = 0$ we get by part (1)

$$
\begin{aligned}
(\nabla_{E_j}\nabla_{[E_i,E_j]}h)(X,E_i) &= (\nabla_{[E_i,E_j]}\nabla_{E_j}h)(X,E_i) + (\nabla_{[E_j,[E_i,E_j]]}h)(X,E_i) \\
&= (\nabla_{\nabla_{E_j}(\nabla_{E_i}E_j-\nabla_{E_j}E_i)}h)(X,E_i)
\end{aligned}
$$

and thus

$$
\begin{aligned}
(\nabla_{E_j}\nabla_{E_i}\nabla_{E_j}h)(X,E_i) &= (\nabla_{E_j}\nabla_{E_j}\nabla_{E_i}h)(X,E_i) + (\nabla_{\nabla_{E_j}(\nabla_{E_i}E_j-\nabla_{E_j}E_i)}h)(X,E_i) \\
&- (\nabla_{E_j}h)(R(E_i,E_j)X,E_i) - h((\nabla_{E_j}R)(E_i,E_j)X,E_i) \\
&- (\nabla_{E_j}h)(X,R(E_i,E_j)E_i) - h(X,(\nabla_{E_j}R)(E_i,E_j)E_i).
\end{aligned}
$$

Inserting this into (17) we get

$$
\begin{aligned}
(\nabla_{E_i}\nabla_{E_j}\nabla_{E_j}h)(X,E_i) &= (\nabla_{E_j}\nabla_{E_j}\nabla_{E_i}h)(X,E_i) + (\nabla_{\nabla_{E_j}(\nabla_{E_i}E_j-\nabla_{E_j}E_i)}h)(X,E_i) \\
&- 2(\nabla_{E_j}h)(R(E_i,E_j)X,E_i) - h((\nabla_{E_j}R)(E_i,E_j)X,E_i) \\
&- 2(\nabla_{E_j}h)(X,R(E_i,E_j)E_i) - 2(X,(\nabla_{E_j}R)(E_i,E_j)E_i)
\end{aligned}
$$

and inserting this into (16) we obtain

$$
\begin{aligned}
(\operatorname{div}\nabla^*\nabla h)(X) = \sum_{i,j=1}^{n} \varepsilon_i\varepsilon_j\{&-(\nabla_{E_j}\nabla_{E_j}\nabla_{E_i}h)(X,E_i) - (\nabla_{\nabla_{E_j}(\nabla_{E_i}E_j-\nabla_{E_j}E_i)}h)(X,E_i) \\
&+ 2(\nabla_{E_j}h)(R(E_i,E_j)X,E_i) + h((\nabla_{E_j}R)(E_i,E_j)X,E_i) \\
&+ 2(\nabla_{E_j}h)(X,R(E_i,E_j)E_i) + 2h(X,(\nabla_{E_j}R)(E_i,E_j)E_i) + (\nabla_{\nabla_{E_i}\nabla_{E_j}E_j}h)(X,E_i)\}.
\end{aligned}
$$

Using (15) we conclude that

$$
\begin{aligned}
(\operatorname{div}\Delta_L h)(X) &= (\operatorname{div}\nabla^*\nabla h)(X) - 2(\operatorname{div}\mathring{R}h)(X) \\
&= \sum_{i,j=1}^{n} \varepsilon_i\varepsilon_j\{-(\nabla_{E_j}\nabla_{E_j}\nabla_{E_i}h)(X,E_i) - (\nabla_{\nabla_{E_j}(\nabla_{E_i}E_j-\nabla_{E_j}E_i)}h)(X,E_i) \\
&+ h((\nabla_{E_j}R)(E_i,E_j)X,E_i) \\
&+ 2(\nabla_{E_j}h)(X,R(E_i,E_j)E_i) + 2h(X,(\nabla_{E_j}R)(E_i,E_j)E_i) + (\nabla_{\nabla_{E_i}\nabla_{E_j}E_j}h)(X,E_i)\} \\
&= \sum_{i,j=1}^{n} \varepsilon_i\varepsilon_j\{-(\nabla_{E_j}\nabla_{E_j}\nabla_{E_i}h)(X,E_i) + (\nabla_{\nabla_{E_j}\nabla_{E_j}E_i}h)(X,E_i)\}
\end{aligned}
$$

$$+\sum_{i,j,k=1}^{n}\varepsilon_i\varepsilon_j\varepsilon_k\{\left\langle R(E_i,E_j)E_j,E_k\right\rangle(\nabla_{E_k}h)(X,E_i)+\left\langle(\nabla_{E_j}R)(E_i,E_j)X,E_k\right\rangle h(E_k,E_i)$$
$$+2\left\langle R(E_i,E_j)E_i,E_k\right\rangle(\nabla_{E_j}h)(X,E_k)+2\left\langle(\nabla_{E_j}R)(E_i,E_j)E_i,E_k\right\rangle h(X,E_k)\}.$$

We note that all triple sums on the right hand side vanish. Namely the first and the third triple sum are proportional to ric and therefore vanish. The fourth triple sum is proportional to derivatives of ric and therefore vanishes as well. Moreover by the first Bianchi identity we have

$$\begin{aligned}\left\langle(\nabla_{E_j}R)(E_i,E_j)X,E_k\right\rangle &= -\left\langle(\nabla_{E_j}R)(X,E_k)E_i,E_j\right\rangle\\ &= \left\langle(\nabla_X R)(E_k,E_j)E_i,E_j\right\rangle+\left\langle(\nabla_{E_k}R)(E_j,X)E_i,E_j\right\rangle.\end{aligned}$$

Thus the second triple sum is proportional to derivatives of ric and therefore vanishes. We conclude that

$$(\operatorname{div}\Delta_L h)(X)=\sum_{i,j=1}^{n}\varepsilon_i\varepsilon_j\{-(\nabla_{E_j}\nabla_{E_j}\nabla_{E_i}h)(X,E_i)+(\nabla_{\nabla_{E_j}\nabla_{E_j}E_i}h)(X,E_i)\}. \tag{18}$$

For all j we have

$$\begin{aligned}(\nabla_{E_j}\operatorname{div}h)(X) &= \partial_{E_j}\operatorname{div}h(X)-\operatorname{div}h(\nabla_{E_j}X)\\ &= \partial_{E_j}\Big(\sum_{i=1}^{n}\varepsilon_i(\nabla_{E_i}h)(X,E_i)\Big)-\sum_{i=1}^{n}\varepsilon_i(\nabla_{E_i}h)(\nabla_{E_j}X,E_i)\\ &= \sum_{i=1}^{n}\varepsilon_i(\nabla_{E_j}\nabla_{E_i}h)(X,E_i)+\sum_{i=1}^{n}\varepsilon_i(\nabla_{E_i}h)(X,\nabla_{E_j}E_i).\end{aligned}$$

Since $\nabla E_j|_p=0$ we have $\nabla^*\nabla\operatorname{div}h=-\sum_{j=1}^{n}\varepsilon_j\nabla_{E_j}\nabla_{E_j}\operatorname{div}h$ and thus

$$\begin{aligned}(\nabla^*\nabla\operatorname{div}h)(X) &= -\sum_{j=1}^{n}\varepsilon_j\partial_{E_j}(\nabla_{E_j}\operatorname{div}h)(X)\\ &= \sum_{i,j=1}^{n}\varepsilon_i\varepsilon_j\partial_{E_j}\{-(\nabla_{E_j}\nabla_{E_i}h)(X,E_i)-(\nabla_{E_i}h)(X,\nabla_{E_j}E_i)\}\\ &= \sum_{i,j=1}^{n}\varepsilon_i\varepsilon_j\{-(\nabla_{E_j}\nabla_{E_j}\nabla_{E_i}h)(X,E_i)-(\nabla_{E_i}h)(X,\nabla_{E_j}\nabla_{E_j}E_i)\}.\end{aligned}$$

Thus by (18) and since $\langle E_i,E_k\rangle$ is constant we get

$$\begin{aligned}(\operatorname{div}\Delta_L h)(X)-(\nabla^*\nabla\operatorname{div}h)(X) &= \sum_{i,j=1}^{n}\varepsilon_i\varepsilon_j\{(\nabla_{\nabla_{E_j}\nabla_{E_j}E_i}h)(X,E_i)+(\nabla_{E_i}h)(X,\nabla_{E_j}\nabla_{E_j}E_i)\}\\ &= \sum_{i,j,k=1}^{n}\varepsilon_i\varepsilon_j\varepsilon_k\{\left\langle\nabla_{E_j}\nabla_{E_j}E_i,E_k\right\rangle+\left\langle E_i,\nabla_{E_j}\nabla_{E_j}E_k\right\rangle\}(\nabla_{E_k}h)(X,E_i)\\ &= \sum_{i,j,k=1}^{n}\varepsilon_i\varepsilon_j\varepsilon_k\{\partial_{E_j}\partial_{E_j}\langle E_i,E_k\rangle\}(\nabla_{E_k}h)(X,E_i)\\ &= 0.\end{aligned}$$

This finishes the proof. □

Let M be a time-oriented Lorentzian manifold with $\mathrm{ric} \equiv 0$. Our aim is to construct a solution to the gravitational wave equation

$$\Delta_L h - \operatorname{div}{}^*\operatorname{div} h - \tfrac{1}{2}\nabla d\,\langle g, h\rangle = 0. \tag{19}$$

Let $\widehat{M} \subset M$ be a spacelike hypersurface. We assume that $\widehat{M}$ is a *Cauchy hypersurface*, i.e. every maximal timelike curve in M meets $\widehat{M}$ exactly once. Let ν be the future directed unit normal vector field along $\widehat{M}$.

Theorem 5.12. *Let* $\mathrm{ric} \equiv 0$ *on* M *and let* $\widehat{M} \subset M$ *be a spacelike Cauchy hypersurface such that* $\widehat{K} \equiv 0$ *on* $\widehat{M}$. *Then for every solution* $(\widehat{h}, \widehat{k})$ *to the linearization* (10), (14) *of the constraint equations at* $(\widehat{g}, \widehat{K})$ *there exists a solution* h *to the gravitational wave equation* (19) *such that for all* $X, Y \in T\widehat{M}$ *we have* $h(X,Y) = \widehat{h}(X,Y)$ *and* $(\nabla_\nu h)(X,Y) = \widehat{k}(X,Y)$ *and such that* h *satisfies* $\operatorname{div} h \equiv 0$ *and* $\langle g, h\rangle \equiv 0$.

Remark 5.13. **(1)** In the physics literature such a solution is called to be in TT-gauge, where TT means transverse and traceless.

(2) The assumptions of the theorem are satisfied e.g. if M is Schwarzschild spacetime with coordinates $(t, r, \sigma) \in \mathbb{R} \times ((0, 2m) \cup (2m, \infty)) \times S^2$ and $\widehat{M} = \{(t, r, \sigma) \in M \,|\, t = t_0\}$, where $t_0 \in \mathbb{R}$ is fixed. $\widehat{M}$ is a totally geodesic hypersurface since it is the fixed point set of the isometry $(t, r, \sigma) \mapsto (2t_0 - t, r, \sigma)$.

Proof. By Theorem 3.2.11 in [?] we know that there exists a unique solution $h \in \Gamma(\odot^2 T^*M)$ to the wave equation $\Delta_L h = 0$ which has the following initial data along $\widehat{M}$:

$$\begin{aligned}
h(X,Y) &= \widehat{h}(X,Y), & (\nabla_\nu h)(X,Y) &= \widehat{k}(X,Y),\\
h(\nu,X) &= (\operatorname{div}\widehat{k})(X), & (\nabla_\nu h)(\nu,X) &= (\operatorname{div}\widehat{h})(X),\\
h(\nu,\nu) &= \left\langle \widehat{g}, \widehat{k}\right\rangle, & (\nabla_\nu h)(\nu,\nu) &= \left\langle \widehat{g}, \widehat{k}\right\rangle,
\end{aligned}$$

where $X, Y \in T\widehat{M}$. We claim that h is a solution to the gravitational wave equation (19). In order to prove this let $(E_i)_{i=1}^{n-1}$ be a local orthonormal frame tangential to $\widehat{M}$. Then along $\widehat{M}$ we get using the initial data

$$\langle g, h\rangle = -h(\nu,\nu) + \sum_{i=1}^{n-1} h(E_i, E_i) = -h(\nu,\nu) + \sum_{i=1}^{n-1} \widehat{h}(E_i, E_i) = -h(\nu,\nu) + \left\langle \widehat{g}, \widehat{h}\right\rangle = 0$$

$$\partial_\nu \langle g, h\rangle = \langle g, \nabla_\nu h\rangle = -(\nabla_\nu h)(\nu,\nu) + \sum_{i=1}^{n-1} (\nabla_\nu h)(E_i, E_i) = -(\nabla_\nu h)(\nu,\nu) + \left\langle \widehat{g}, \widehat{k}\right\rangle = 0.$$

Using Lemma 5.4 we conclude that the function $\langle g, h\rangle$ is a solution to the wave equation $\nabla^*\nabla\langle g, h\rangle = \langle g, \Delta_L h\rangle = 0$ on M with the initial data $\langle g, h\rangle = \partial_\nu\langle g, h\rangle = 0$ along $\widehat{M}$. By uniqueness of the solution to the wave equation with these initial data it follows that $\langle g, h\rangle \equiv 0$ on M. Next we show that $\operatorname{div} h \equiv 0$ on M. Let $X \in T_p\widehat{M}$, $p \in \widehat{M}$, and extend X to a locally defined vector field such that at p we have $\nabla X|_p = 0$. We also assume that at p we have $\nabla E_i|_p = 0$, $i = 1, ..., n-1$. Since $\widehat{K} \equiv 0$ by assumption, we conclude that we also have $\widehat{\nabla} E_i|_p = \widehat{\nabla} X|_p = 0$. Thus along $\widehat{M}$ we get using the initial data

$$\begin{aligned}
(\operatorname{div} h)(X) &= -(\nabla_\nu h)(\nu, X) + \sum_{i=1}^{n-1}(\nabla_{E_i} h)(E_i, X)\\
&= -(\nabla_\nu h)(\nu, X) + \sum_{i=1}^{n-1}\partial_{E_i} h(E_i, X)\\
&= -(\nabla_\nu h)(\nu, X) + \sum_{i=1}^{n-1}\partial_{E_i}\widehat{h}(E_i, X)\\
&= -(\nabla_\nu h)(\nu, X) + \sum_{i=1}^{n-1}((\widehat{\nabla}_{E_i}\widehat{h})(E_i, X) + \widehat{h}(\widehat{\nabla}_{E_i}E_i, X) + \widehat{h}(E_i, \widehat{\nabla}_{E_i}X))\\
&= -(\nabla_\nu h)(\nu, X) + (\operatorname{div}\widehat{h})(X)\\
&= 0.
\end{aligned}$$

Since $\operatorname{ric} \equiv 0$ and $\widehat{K} \equiv 0$ we have $\widehat{\operatorname{ric}} \equiv 0$ and thus by (10), (14) we get

$$\begin{aligned}
0 &= \Delta\left\langle \widehat{g}, \widehat{h}\right\rangle - \delta\operatorname{div}\widehat{h},\\
0 &= \operatorname{div}\widehat{k} - d\left\langle \widehat{g}, \widehat{k}\right\rangle.
\end{aligned}$$

Using the above initial data we get along $\widehat{M}$

$$\begin{aligned}
(\operatorname{div} h)(\nu) &= -(\nabla_\nu h)(\nu, \nu) + \sum_{i=1}^{n-1}(\nabla_{E_i} h)(E_i, \nu)\\
&= -(\nabla_\nu h)(\nu, \nu) + \sum_{i=1}^{n-1}\partial_{E_i} h(E_i, \nu)\\
&= -(\nabla_\nu h)(\nu, \nu) + \sum_{i=1}^{n-1}\partial_{E_i}((\operatorname{div}\widehat{k})(E_i))\\
&= -\left\langle \widehat{g}, \widehat{k}\right\rangle - \delta\operatorname{div}\widehat{k}.
\end{aligned}$$

By some computations we obtain for all $X \in T\widehat{M}$

$$(\operatorname{div} h)(\nu) = 0, \quad (\nabla_\nu\operatorname{div} h)(X) = 0, \quad (\nabla_\nu\operatorname{div} h)(\nu) = 0.$$

Thus by Lemma 5.11 we conclude that $\operatorname{div} h$ satisfies the wave equation $\nabla^*\nabla\operatorname{div} h = \operatorname{div}(\Delta_L h) = 0$ with initial conditions $\operatorname{div} h = 0$, $\nabla_\nu\operatorname{div} h = 0$ along $\widehat{M}$. By uniqueness of the solution to the wave equation with these initial data it follows that $\operatorname{div} h \equiv 0$ on M. □

5.5. Construction of solutions to the constraint equations

Let $\widehat{M}$ be a 3-dimensional manifold. Our aim is to find Riemannian metrics $\widehat{g}$ and symmetric $(0,2)$-tensor fields $\widehat{K}$ on $\widehat{M}$ which satisfy the constraint equations

$$\begin{aligned}\widehat{\mathrm{scal}} + \left\langle \widehat{g}, \widehat{K} \right\rangle^2 - |\widehat{K}|^2_{\widehat{g}} &= 0,\\ \operatorname{div} \widehat{K} - d\left\langle \widehat{g}, \widehat{K} \right\rangle &= 0.\end{aligned}$$

The space of solutions is not understood in full generality. We introduce the so-called conformal method which can be used to construct solutions in certain cases.

Definition 5.14. Let $(\widehat{M}, \widehat{g})$ be a Riemannian manifold of dimension n. The operator

$$\begin{aligned} L: \quad \Omega^1(\widehat{M}) \to \Gamma(\odot^2 T^* \widehat{M}), \quad \omega \mapsto L\omega := -2\operatorname{div}^* \omega + \frac{2}{n}\delta\omega \cdot \widehat{g}, \text{ i.e.}\\ L\omega(X,Y) = (\nabla_X \omega)(Y) + (\nabla_Y \omega)(X) - \frac{2}{n}\left\langle \nabla\omega, \widehat{g} \right\rangle \widehat{g}(X,Y)\end{aligned}$$

is called the *conformal Killing operator*.

Remark 5.15. **(1)** If V is a Killing vector field, i.e. ∇V is skew-symmetric, then $\omega := V^\flat$ satisfies $L\omega = 0$.

(2) We have $L\omega = 0$ if and only if $V := \omega^\sharp$ is a conformal Killing vector field. Here a vector field V on $\widehat{M}$ is called a conformal Killing vector field of $(\widehat{M}, \widehat{g})$ if the flow $(\Phi_s)_s$ of V consists of conformal diffeomorphisms of $(\widehat{M}, \widehat{g})$, i.e. $\Phi_s^* \widehat{g}$ is conformally equivalent to $\widehat{g}$ for all s.

(3) $L\omega$ is trace-free since we have

$$\left\langle L\omega, \widehat{g} \right\rangle = 2\sum_{i=1}^n (\nabla_{E_i}\omega)(E_i) - \frac{2}{n}\left\langle \nabla\omega, \widehat{g} \right\rangle \underbrace{\left\langle \widehat{g}, \widehat{g} \right\rangle}_{=n} = 0.$$

From now we take $\dim \widehat{M} = 3$ and we assume that the following quantities are given on $\widehat{M}$:

g_0, a Riemannian metric,

σ, a symmetric $(0,2)$-tensor field which is trace-free and divergence-free with respect to g_0, i.e. $\langle g_0, \sigma \rangle = 0$ and $\operatorname{div}_{g_0}\sigma = 0$,

τ, a function on $\widehat{M}$.

We determine on $\widehat{M}$:

Φ, a positive function,

ω, a 1-form

by solving the following system of equations if this is possible

$$\begin{aligned} \operatorname{div}_0 L_0\omega &= \frac{2}{3}\Phi^6 d\tau, \\ -\Delta_0\Phi &= \frac{1}{8}\mathrm{scal}_0\Phi - \frac{1}{8}|\sigma + L_0\omega|_0^2\,\Phi^{-7} + \frac{1}{12}\tau^2\Phi^5, \end{aligned} \tag{20}$$

where the subscripts "0" indicate that the respective quantities are defined with respect to g_0. In the literature the second equation is often called the *Lichnerowicz equation*.

Remark 5.16. The operator $\operatorname{div}_0 \circ L_0\colon \Omega^1(\widehat{M}) \to \Omega^1(\widehat{M})$ is a linear elliptic differential operator of second order. Thus (20) is a semilinear elliptic system of equations and is thus more convenient for analysis than the original constraint equations.

Theorem 5.17. *Assume that* (Φ, ω) *solve the system* (20). *Then*

$$\widehat{g} := \Phi^4 g_0, \quad \widehat{K} := \Phi^{-2}(\sigma + L_0\omega) + \frac{1}{3}\tau\Phi^4 g_0$$

solve the constraint equations.

Remark 5.18. **(1)** This method of constructing solutions to the constraint equations is called the *conformal method* since we have fixed the conformal class of $\widehat{g}$ by fixing g_0.

(2) The constraint equations form an underdetermined system of equations. After fixing τ, σ and the conformal class of $\widehat{g}$ we are left to solve the system (20) which is no longer underdetermined.

(3) In local coordinates we have $\widehat{g}^{ij} = \Phi^{-4}g_0^{ij}$. Thus for $(0,2)$-tensor fields μ, λ on $\widehat{M}$ we get

$$\langle \mu, \lambda\rangle_{\widehat{g}} = \mu_{ij}\lambda_{k\ell}\widehat{g}^{ik}\widehat{g}^{j\ell} = \Phi^{-8}\mu_{ij}\lambda_{k\ell}g_0^{ik}g_0^{j\ell} = \Phi^{-8}\langle \mu, \lambda\rangle_0\,. \tag{21}$$

It follows that

$$\left\langle \widehat{g}, \widehat{K}\right\rangle_{\widehat{g}} = \Phi^{-6}\langle g_0, \sigma + L_0\omega\rangle_{g_0} + \tau = \tau$$

and therefore τ is the mean curvature of $\widehat{M}$ in the Einstein development of M.

For every Riemannian metric g on $\widehat{M}$ the *conformal Laplace operator* or *Yamabe operator* for g is defined by

$$Y_g := \Delta_g + \frac{1}{8}\mathrm{scal}_g.$$

Lemma 5.19. *Let* $\dim \widehat{M} = 3$ *and let* $\widehat{g} = \Phi^4 g_0$. *Then for all* $u \in C^\infty(\widehat{M})$ *we have*

$$Y_{\widehat{g}}(\Phi^{-1}u) = \Phi^{-5}Y_0(u).$$

Proof. The proof can be found in [**?**], Folgerung 2.1.5. □

Using this lemma with $u = \Phi$ we get

$$\frac{1}{8}\mathrm{scal}_{\widehat{g}} = Y_{\widehat{g}}(1) = \Phi^{-5}Y_0(\Phi) = \Phi^{-5}\Big(\Delta_0\Phi + \frac{1}{8}\mathrm{scal}_0\Phi\Big)$$

and therefore

$$\mathrm{scal}_{\widehat{g}} = \Phi^{-5}(8\Delta_0\Phi + \mathrm{scal}_0\Phi). \tag{22}$$

Lemma 5.20. *Let* $\dim \widehat{M} = 3$ *and let* $\widehat{g} = \Phi^4 g_0$*. For all trace-free symmetric* $(0,2)$*-tensor fields* μ *on* $\widehat{M}$ *we have*

$$\operatorname{div}_{\widehat{g}}\mu = \Phi^{-4}\{\operatorname{div}_0\mu + 2\Phi^{-1}\mu(\nabla^0\Phi,\cdot)\}.$$

Proof. We fix a point $p \in \widehat{M}$. Let $(E_i)_{i=1}^3$ be a local g_0-orthonormal frame of $T\widehat{M}$ with respect to g_0 such that at p we have $\nabla^0 E_i|_p = 0$ for all i. Let $X \in T_p\widehat{M}$ and extend X to a vector field defined on an open neighborhood of p such that at p we have $\nabla^0 X|_p = 0$. Then the vectors $\widehat{E_i} := \Phi^{-2}E_i$, $i = 1,2,3$ form a local $\widehat{g}$-orthonormal frame. We compute

$$\begin{aligned}(\operatorname{div}_{\widehat{g}}\mu)(X) &= \sum_{i=1}^3(\widehat{\nabla}_{\widehat{E_i}}\mu)(\widehat{E_i},X)\\ &= \Phi^{-4}\sum_{i=1}^3(\widehat{\nabla}_{E_i}\mu)(E_i,X)\\ &= \Phi^{-4}\sum_{i=1}^3(\partial_{E_i}\mu(E_i,X) - \mu(\widehat{\nabla}_{E_i}E_i,X) - \mu(E_i,\widehat{\nabla}_{E_i}X)).\end{aligned}$$

We write $\Phi^4 = e^{2f}$ where $f \in C^\infty(\widehat{M})$ and we note that the Levi-Civita connections $\widehat{\nabla}$, ∇^0 for $\widehat{g}$ and g_0 are related by the formula (see [**?**], Lemma 2.1.2)

$$\widehat{\nabla}_X Y = \nabla^0_X Y + df(X)Y + df(Y)X - g_0(X,Y)\nabla^0 f.$$

We conclude that

$$\begin{aligned}(\operatorname{div}_{\widehat{g}}\mu)(X) = \Phi^{-4}\Big\{&(\operatorname{div}_0\mu)(X) - \sum_{i=1}^3\mu(2df(E_i)E_i - g_0(E_i,E_i)\nabla^0 f, X)\\ &- \sum_{i=1}^3\mu(E_i, df(E_i)X + df(X)E_i - g_0(X,E_i)\nabla^0 f)\Big\}\\ = \Phi^{-4}\{&(\operatorname{div}_0\mu)(X) - 2\mu(\nabla^0 f, X) + 3\mu(\nabla^0 f, X)\end{aligned}$$

$$
\begin{aligned}
&- \mu(\nabla^0 f, X) - df(X)\langle \mu, g_0\rangle_{g_0} + \mu(X, \nabla^0 f)\} \\
&= \Phi^{-4}\{(\mathrm{div}_0 \mu)(X) + \mu(X, \nabla^0 f)\}.
\end{aligned}
$$

Since we have $f = 2\ln\Phi$ we get $\nabla^0 f = 2\Phi^{-1}\nabla^0\Phi$. This finishes the proof. □

Proof of Theorem 5.17. Using (21) we compute

$$
|\widehat{K}|^2_{\widehat{g}} = \left|\Phi^{-2}(\sigma + L_0\omega) + \frac{1}{3}\tau\widehat{g}\right|^2_{\widehat{g}} = \Phi^{-4}|\sigma + L_0\omega|^2_{\widehat{g}} + \frac{\tau^2}{9}|\widehat{g}|^2_{\widehat{g}} = \Phi^{-12}|\sigma + L_0\omega|^2_0 + \frac{\tau^2}{3}
$$

and therefore using (22) we obtain

$$
\begin{aligned}
\widehat{\mathrm{scal}} + \left\langle \widehat{g}, \widehat{K}\right\rangle^2_{\widehat{g}} - |\widehat{K}|^2_{\widehat{g}} &= \Phi^{-5}(8\Delta_0\Phi + \mathrm{scal}_0\Phi) + \tau^2 - \left(\Phi^{-12}|\sigma + L_0\omega|^2_0 + \frac{\tau^2}{3}\right) \\
&= 8\Phi^{-5}\left(\Delta_0\Phi + \frac{1}{8}\mathrm{scal}_0\Phi - \frac{1}{8}|\sigma + L_0\omega|^2_0\Phi^{-7} + \frac{1}{12}\tau^2\Phi^5\right) \\
&= 0.
\end{aligned}
$$

Thus the Hamiltonian constraint is satisfied. Let $(\widehat{E}_i)_{i=1}^3$ be a local $\widehat{g}$-orthonormal frame of $T\widehat{M}$. We note that for all $f \in C^\infty(\widehat{M})$ and for all symmetric $(0,2)$-tensor fields μ on $\widehat{M}$ we have

$$
\mathrm{div}_{\widehat{g}}(f\mu) = \sum_{i=1}^{3} \widehat{\nabla}_{\widehat{E}_i}(f\mu)(\widehat{E}_i, \cdot) = \sum_{i=1}^{3}(\partial_{\widehat{E}_i} f \cdot \mu + f\widehat{\nabla}_{\widehat{E}_i}\mu)(\widehat{E}_i, \cdot) = \mu(\widehat{\nabla} f, \cdot) + f\,\mathrm{div}_{\widehat{g}}\mu.
$$

Using this and Lemma 5.20 we compute

$$
\begin{aligned}
\mathrm{div}_{\widehat{g}}\widehat{K} &= \mathrm{div}_{\widehat{g}}\left(\Phi^{-2}(\sigma + L_0\omega) + \frac{\tau}{3}\widehat{g}\right) \\
&= (\sigma + L_0\omega)(\widehat{\nabla}\Phi^{-2}, \cdot) + \Phi^{-2}\mathrm{div}_{\widehat{g}}(\sigma + L_0\omega) + \frac{1}{3}\widehat{g}(\widehat{\nabla}\tau, \cdot) \\
&= -2\Phi^{-3}(\sigma + L_0\omega)(\widehat{\nabla}\Phi, \cdot) + \Phi^{-6}\mathrm{div}_0(\sigma + L_0\omega) + 2\Phi^{-7}(\sigma + L_0\omega)(\nabla^0\Phi, \cdot) + \frac{1}{3}d\tau.
\end{aligned}
$$

Since for all $u \in C^\infty(\widehat{M})$ we have

$$
\widehat{\nabla} u = \widehat{g}^{ij}\partial_i u \cdot \partial_j = \Phi^{-4} g_0^{ij}\partial_i u \cdot \partial_j = \Phi^{-4}\nabla^0 u
$$

we obtain

$$
\begin{aligned}
\mathrm{div}_{\widehat{g}}\widehat{K} &= -2\Phi^{-7}(\sigma + L_0\omega)(\nabla^0\Phi, \cdot) + \Phi^{-6}\mathrm{div}_0(\sigma + L_0\omega) + 2\Phi^{-7}(\sigma + L_0\omega)(\nabla^0\Phi, \cdot) + \frac{1}{3}d\tau \\
&= \Phi^{-6}\mathrm{div}_0(\sigma + L_0\omega) + \frac{1}{3}d\tau \\
&= \Phi^{-6}\mathrm{div}_0(L_0\omega) + \frac{1}{3}d\tau \\
&= \frac{2}{3}d\tau + \frac{1}{3}d\tau \\
&= d\left\langle \widehat{g}, \widehat{K}\right\rangle_{\widehat{g}}.
\end{aligned}
$$

Thus also the momentum constraint is satisfied. □

Now the question arises whether the equations (20) can be solved. The following example shows that this is unfortunately not always the case.

Example 5.21. Let $\widehat{M}$ be compact without boundary, let g_0 be a Riemannian metric on $\widehat{M}$ such that $\mathrm{scal}_0 \geq 0$, $\mathrm{scal}_0 \not\equiv 0$. Let $\sigma \equiv 0$ and let τ be constant. Assume that (Φ, ω) is a solution to (20). Then we have $\mathrm{div}_0 L_0\omega = 0$ and thus by the second equation

$$\begin{aligned} 0 &= \int_{\widehat{M}} \langle \mathrm{div}_0 L_0\omega, \omega\rangle \, dv = \int_{\widehat{M}} \langle L_0\omega, \mathrm{div}^*\omega\rangle \, dv = \int_{\widehat{M}} \left\langle L_0\omega, -\tfrac{1}{2}L_0\omega + \frac{1}{3}\delta\omega \cdot g_0 \right\rangle dv \\ &= -\tfrac{1}{2}\int_{\widehat{M}} |L_0\omega|_0^2 dv + \frac{1}{3}\int_{\widehat{M}} \delta\omega \underbrace{\langle L_0\omega, g_0\rangle}_{=0} dv = -\tfrac{1}{2}\|L_0\omega\|^2_{L^2(\widehat{M})}. \end{aligned}$$

We conclude that $L_0\omega = 0$. Substituting this into the first equation we get

$$-\Delta_0\Phi = \frac{1}{8}\mathrm{scal}_0\Phi + \frac{1}{12}\tau^2\Phi^5$$

and thus

$$0 = -\int_{\widehat{M}} \Delta_0\Phi \, dv = \int_{\widehat{M}} \left(\frac{1}{8}\mathrm{scal}_0\Phi + \frac{\tau^2}{12}\Phi^5\right) dv.$$

Since $\mathrm{scal}_0 \geq 0$ and $\mathrm{scal}_0 \not\equiv 0$ and $\Phi > 0$ the right hand side is strictly positive which is a contradiction. The same argument also shows that there is no solution to (20) if

- $\mathrm{scal}_0 \geq 0$, $\tau \equiv \mathrm{const} \neq 0$, $\sigma \equiv 0$,
- $\mathrm{scal}_0 \equiv 0$, $\tau \equiv 0$, $\sigma \not\equiv 0$
- $\mathrm{scal}_0 < 0$, $\tau \equiv 0$.

Next we construct solutions to the system (20) in some cases using the method of sub- and supersolutions as in the article [?] by Isenberg. The following theorem shows how this method works.

Theorem 5.22. *Let $\widehat{M}$ be a closed Riemannian manifold of dimension n. Let $f \in C^\infty(\widehat{M} \times (0,\infty))$ and let $\Phi_-, \Phi_+ \in C^\infty(\widehat{M})$ such that*

$$\begin{aligned} 0 < \Phi_- &\leq \Phi_+, \\ -\Delta\Phi_- &\geq f(x, \Phi_-), \\ -\Delta\Phi_+ &\leq f(x, \Phi_+), \end{aligned}$$

where $f(x, \Phi_\pm)$ denote the functions $x \mapsto f(x, \Phi_\pm(x))$. Then there exists $\Phi \in C^\infty(\widehat{M})$ such that

$$\begin{aligned} \Phi_- \leq \Phi &\leq \Phi_+, \\ -\Delta\Phi &= f(x, \Phi). \end{aligned}$$

Remark 5.23. For every $\rho > 0$ the operator $P := \Delta + \rho$ is strictly positive and thus invertible on $L^2(\widehat{M})$. For P the following *maximum principle* holds: If $P\Phi \geq 0$ on $\widehat{M}$ then $\Phi \geq 0$ on $\widehat{M}$. Namely if $k(t, x, y)$ is the heat kernel of Δ on $\widehat{M}$, then $e^{-\rho t}k(t, x, y)$ is the heat kernel of P, i.e. the integral kernel of the operator e^{-tP}. We conclude that the integral kernel of P^{-1} (i.e. the Green function of P) is given by

$$G(x, y) = \int_0^\infty e^{-\rho t}k(t, x, y)dt \in (0, \infty], \quad \text{for all } x, y \in \widehat{M}$$

since we have

$$\int_0^\infty e^{-tP}dt = -P^{-1}e^{-tP}|_{t=0}^\infty = P^{-1}.$$

Therefore if $P\Phi \geq 0$, then we get for all $x \in \widehat{M}$

$$\Phi(x) = P^{-1}P\Phi(x) = \int_{\widehat{M}} \underbrace{G(x, y)}_{>0} \underbrace{(P\Phi)(y)}_{\geq 0} dy \geq 0.$$

Proof of Theorem 5.22. (a) We put

$$I := [\min_{\widehat{M}} \Phi_-, \max_{\widehat{M}} \Phi_+].$$

Then $\widehat{M} \times I$ is compact and thus there exists $\rho > 0$ such that for all $x \in \widehat{M}$ and for all $s \in I$ we have $\frac{\partial f}{\partial s}(x, s) \leq \rho$. We put $P := \Delta + \rho$ and we define

$$F : \quad \widehat{M} \times I \to \mathbb{R}, \quad F(x, s) := \rho s - f(x, s).$$

Then we have $\frac{\partial F}{\partial s} = \rho - \frac{\partial f}{\partial s} \geq 0$ on $\widehat{M} \times I$. The hypotheses on $\Phi_\pm$ read

$$\begin{aligned} -\Delta\Phi_- \geq f(x, \Phi_-) &\Longleftrightarrow P\Phi_- \leq F(x, \Phi_-), \\ -\Delta\Phi_+ \leq f(x, \Phi_+) &\Longleftrightarrow P\Phi_+ \geq F(x, \Phi_+) \end{aligned}$$

and we have

$$-\Delta\Phi = f(x, \Phi) \Longleftrightarrow P\Phi = F(x, \Phi).$$

(b) We solve inductively

$$P\Phi_{j+1} = F(x, \Phi_j), \quad j = 0, 1, 2, \dots, \quad \Phi_0 = \Phi_+.$$

By elliptic regularity we know that $\Phi_j \in C^\infty(\widehat{M})$ for all j. We claim that for all j we have

$$\Phi_j \geq \Phi_{j+1} \geq \Phi_-.$$

In order to show the claim for $j = 0$ we use that

$$P(\Phi_+ - \Phi_1) = P\Phi_+ - P\Phi_1 \geq F(x, \Phi_+) - F(x, \Phi_+) = 0.$$

By the maximum principle we get $\Phi_+ - \Phi_1 \geq 0$, i.e. $\Phi_1 \leq \Phi_+$. Furthermore we have

$$P(\Phi_1 - \Phi_-) \geq F(x, \Phi_+) - F(x, \Phi_-) \geq 0,$$

since $\Phi_- \leq \Phi_+$ and $s \mapsto F(x, s)$ is monotonically increasing on I. By the maximum principle we conclude $\Phi_1 - \Phi_- \geq 0$, i.e. $\Phi_- \leq \Phi_1$. This shows the claim for $j = 0$. Now assume that the claim is true for some $j \geq 1$. We have

$$P(\Phi_j - \Phi_{j+1}) = F(x, \Phi_{j-1}) - F(x, \Phi_j) \geq 0,$$

since $\Phi_{j-1} \geq \Phi_j$ by the inductive hypothesis and since $s \mapsto F(x, s)$ is monotonically increasing on $[\min_{\widehat{M}} \Phi_j, \max_{\widehat{M}} \Phi_{j-1}] \subset I$. By the maximum principle we get $\Phi_j \geq \Phi_{j+1}$ and analogously we show that $\Phi_{j+1} \geq \Phi_-$. We have proved the claim and thus we get

$$\Phi_+ \geq \Phi_1 \geq \Phi_2 \geq \Phi_3 \geq ... \geq \Phi_-.$$

In particular, the sequence $(\Phi_j)_j$ converges pointwise to a function $\Phi\colon \widehat{M} \to \mathbb{R}$ such that we have $\Phi_- \leq \Phi \leq \Phi_+$.
(c) Next we show that $\Phi \in C^\infty(\widehat{M})$. In the following estimates we will use some constants $a_1, a_2, a_3, ...$, which will all be independent of j. First we note that

$$\|F(x, \Phi_j)\|_{C^0(\widehat{M})} \leq \|F\|_{C^0(\widehat{M}\times I)} =: a_1$$

and thus for all $p \geq 1$ we get

$$\|F(x, \Phi_j)\|_{L^p(\widehat{M})} \leq a_1 \mathrm{vol}(\widehat{M})^{1/p} =: a_2.$$

We choose $p > n = \dim \widehat{M}$ and since $P\Phi_{j+1} = F(x, \Phi_j)$ we get by elliptic estimates for Sobolev spaces that there exist $a_3, a_4 > 0$ such that for all j we have

$$\|\Phi_{j+1}\|_{W^{2,p}(\widehat{M})} \leq a_3\|F(x, \Phi_j)\|_{L^p(\widehat{M})} + a_4\|\Phi_{j+1}\|_{L^1(\widehat{M})} \leq a_3 \cdot a_2 + a_4\|\Phi_+\|_{L^1(\widehat{M})} =: a_5.$$

By the Sobolev embedding theorem we have a continuous embedding $W^{2,p}(\widehat{M}) \hookrightarrow C^{0,\alpha}(\widehat{M})$ for some $\alpha > 0$. Thus there exists $a_6 > 0$ such that for all j we have

$$\|\Phi_{j+1}\|_{C^{0,\alpha}(\widehat{M})} \leq a_6$$

and thus there exists $a_7 > 0$ such that for all j we have

$$\|F(x, \Phi_{j+1})\|_{C^{0,\alpha}(\widehat{M})} \leq a_7.$$

By Schauder estimates there exist $a_8, a_9 > 0$ such that for all j we have

$$\|\Phi_{j+1}\|_{C^{2,\alpha}(\widehat{M})} \leq a_8\|F(x, \Phi_j)\|_{C^{0,\alpha}(\widehat{M})} + a_9\|\Phi_j\|_{C^0(\widehat{M})} \leq a_8 \cdot a_7 + a_9\|\Phi_+\|_{C^0(\widehat{M})} =: a_{10}.$$

Iterating this procedure we obtain for every $k \in \mathbb{N}$ a number $b_k > 0$ such that for all j we have $\|\Phi_j\|_{C^{k,\alpha}(\widehat{M})} \leq b_k$. Now we choose α' such that $0 < \alpha' < \alpha$. Then the embedding $C^{k,\alpha}(\widehat{M}) \hookrightarrow C^{k,\alpha'}(\widehat{M})$ is compact and thus after passing to a subsequence $(\Phi_j)_j$ converges in $C^{k,\alpha'}(\widehat{M})$. The limit is the function Φ obtained above. We choose a diagonal subsequence and we get that $\Phi \in C^{k,\alpha'}(\widehat{M})$ for all k, i.e. $\Phi \in C^\infty(\widehat{M})$.
(d) For all j we have $P\Phi_{j+1} = F(x, \Phi_j)$. As $j \to \infty$ the left hand side tends to $P\Phi$ and the right hand side tends to $F(x, \Phi)$. Thus we have $P\Phi = F(x, \Phi)$. □

Now we apply Theorem 5.22 to construct solutions to the system (20) under the following assumptions. Let $(\widehat{M}, g_0)$ be a 3-dimensional closed Riemannian manifold with $\mathrm{scal}_0 < 0$. Let σ be a symmetric $(0,2)$-tensor field on $\widehat{M}$ such that $\langle g_0, \sigma\rangle_0 = 0$ and $\mathrm{div}_0\sigma = 0$ and let τ be a non-zero constant function on $\widehat{M}$. We put $\omega := 0$. Then we have $\mathrm{div}_0 L_0\omega = 0 = \frac{2}{3}\Phi^6 d\tau$ for all $\Phi \in C^\infty(\widehat{M})$. Thus the first equation in (20) is satisfied and it remains to solve the Lichnerowicz equation. We put

$$f : \widehat{M} \times (0,\infty) \to \mathbb{R}, \quad f(x,s) := \frac{1}{8}\mathrm{scal}_0(x)s - \frac{1}{8}|\sigma(x)|_0^2\, s^{-7} + \frac{\tau^2}{12}s^5.$$

We estimate

$$f(x,s) \le \Big(\frac{1}{8}\max_{\widehat{M}}\mathrm{scal}_0 + \frac{\tau^2}{12}s^4\Big)s.$$

Since $\widehat{M}$ is closed and $\mathrm{scal}_0 < 0$ we have $\max_{\widehat{M}}\mathrm{scal}_0 < 0$. Thus there exists $s_- > 0$ such that for all s with $0 < s \le s_-$ and for all $x \in \widehat{M}$ we have $f(x,s) \le 0$. We estimate

$$f(x,s) \ge \Big(\frac{1}{8}\min_{\widehat{M}}\mathrm{scal}_0 s^{-4} - \frac{1}{8}\max_{\widehat{M}}|\sigma|_0^2 s^{-12} + \frac{\tau^2}{12}\Big)s^5.$$

Since $\tau^2 > 0$ there exists $s_+ > 0$ such that for all $s \ge s_+$ and for all $x \in \widehat{M}$ we have $f(x,s) \ge 0$. We put

$$\Phi_- := s_-, \quad \Phi_+ := s_+.$$

Then we have $0 < \Phi_- \le \Phi_+$ and

$$-\Delta_0\Phi_- = 0 \ge f(x,\Phi_-),$$
$$-\Delta_0\Phi_+ = 0 \le f(x,\Phi_+).$$

By Theorem 5.22 there exists $\Phi \in C^\infty(\widehat{M})$ such that $\Phi_- \le \Phi \le \Phi_+$, in particular $\Phi > 0$, and $-\Delta_0\Phi = f(x,\Phi)$, i.e. we get a solution to the Lichnerowicz equation. By Theorem 5.17 we get a solution to the constraint equations.

Remark 5.24. Every closed manifold of dimension 3 carries a Riemannian metric with $\mathrm{scal} < 0$. Thus on every closed manifold of dimension 3 there exist solutions to the constraint equations.

6. Petrov classification

The Petrov classification will divide relativistic spacetimes into types according to algebraic properties of their curvature tensors. So we start with some algebraic considerations.

6.1. Algebraic curvature tensors

Definition 6.1. Let V be a finite dimensional real vector space. A $(0, 4)$-tensor R on V, i.e. a multilinear map

$$R: \quad V \times V \times V \times V \to \mathbb{R}$$

is called an *algebraic curvature tensor* if the following holds for all $X, Y, W, Z \in V$:

(1) skew symmetry: $R(X, Y, W, Z) = -R(Y, X, W, Z) = -R(X, Y, Z, W)$;
(2) Bianchi identity: $R(X, Y, W, Z) + R(X, W, Z, Y) + R(X, Z, Y, W) = 0$.

Lemma 6.2. *If R is an algebraic curvature tensor, then for all $X, Y, W, Z \in V$ we have*

$$R(X, Y, W, Z) = R(W, Z, X, Y).$$

Proof. Using the skew symmetry and the Bianchi identity we get

$$\begin{aligned}
R(X, Y, W, Z) &= -R(X, W, Z, Y) - R(X, Z, Y, W) \\
&= R(W, X, Z, Y) + R(Z, X, Y, W) \\
&= -R(W, Y, X, Z) - R(W, Z, Y, X) - R(Z, W, X, Y) - R(Z, Y, W, X) \\
&= 2R(W, Z, X, Y) + R(Y, W, X, Z) + R(Y, Z, W, X) \\
&= 2R(W, Z, X, Y) - R(Y, X, Z, W) \\
&= 2R(W, Z, X, Y) - R(X, Y, W, Z).
\end{aligned}$$

Adding $R(X, Y, W, Z)$ to both sides and dividing by 2 concludes the proof. □

Now we assume that V carries an inner product $\langle \cdot, \cdot \rangle$ which is non-degenerate but possibly indefinite.

Definition 6.3. The *contraction* of the algebraic curvature tensor R with respect to $\langle\cdot,\cdot\rangle$ is the $(0,2)$-tensor on V defined by

$$C(R)(X,Y) := \sum_{i=1}^{n} \varepsilon_i R(E_i, X, E_i, Y)$$

for $X, Y \in V$, where $(E_i)_{i=1}^n$ is an orthonormal basis of $(V, \langle\cdot,\cdot\rangle)$ and $\varepsilon_i = \langle E_i, E_i\rangle$, $i = 1, ..., n$.

Fact. We recall the universal property of $\Lambda^2(V)$: For every skew-symmetric bilinear map $\beta : V \times V \to W$ there exists a unique linear map $\ell\colon \Lambda^2(V) \to W$ such that $\beta(X,Y) = \ell(X \wedge Y)$ for all $X, Y \in V$.

$$\begin{tikzcd} V \times V \arrow[r, "\wedge"] \arrow[dr, "\forall\beta"'] & \Lambda^2(V) \arrow[d, dotted, "\exists!\ell"] \\ & W \end{tikzcd}$$

Thus for fixed $W, Z \in V$ there is a unique linear map $\tilde{R}^{W,Z}\colon \Lambda^2(V) \to \mathbb{R}$ such that

$$\tilde{R}^{W,Z}(X \wedge Y) = R(X, Y, W, Z)$$

for all $X, Y \in V$. Moreover for fixed $X, Y \in V$ there is a unique linear map $\hat{R}^{X\wedge Y}\colon \Lambda^2(V) \to \mathbb{R}$ such that

$$\hat{R}^{X\wedge Y}(W \wedge Z) = \tilde{R}^{W,Z}(X \wedge Y) = R(X, Y, W, Z)$$

for all $W, Z \in V$. We put with slight abuse of notation

$$R : \Lambda^2(V) \times \Lambda^2(V) \to \mathbb{R}, \quad R(X \wedge Y, W \wedge Z) := \hat{R}^{X\wedge Y}(W \wedge Z) = R(X, Y, W, Z).$$

Note that this map is bilinear and symmetric by Lemma 6.2. The inner product $\langle\cdot,\cdot\rangle$ on V induces an inner product $\langle\cdot,\cdot\rangle$ on $\Lambda^2(V)$ as follows. For $X, Y \in V$ we define

$$\langle X \wedge Y, W \wedge Z\rangle := \langle X, W\rangle \langle Y, Z\rangle - \langle Y, W\rangle \langle X, Z\rangle = \det \begin{pmatrix} \langle X, W\rangle & \langle X, Z\rangle \\ \langle Y, W\rangle & \langle Y, Z\rangle \end{pmatrix}$$

and we extend this definition bilinearly to arbitrary elements of $\Lambda^2(V)$. If $(E_i)_{i=1}^n$ is an orthonormal basis of $(V, \langle\cdot,\cdot\rangle)$ and $\varepsilon_i := \langle E_i, E_i\rangle$, $i = 1, ..., n$, then $(E_i \wedge E_j)_{1\le i<j\le n}$ is an orthonormal basis of $(\Lambda^2(V), \langle\cdot,\cdot\rangle)$. Namely, for $i, j, i', j' \in \{1, ..., n\}$ with $i < j$ and $i' < j'$ and we compute

$$\left\langle E_i \wedge E_j, E_{i'} \wedge E_{j'}\right\rangle = \langle E_i, E_{i'}\rangle \left\langle E_j, E_{j'}\right\rangle - \left\langle E_j, E_{i'}\right\rangle \left\langle E_i, E_{j'}\right\rangle = \delta_{ii'}\varepsilon_i\delta_{jj'}\varepsilon_j - \delta_{ji'}\varepsilon_j\delta_{ij'}\varepsilon_i.$$

If the second term on the right hand side were nonzero we would have $j = i' < j' = i$ contradicting the hypothesis. Thus we have

$$\left\langle E_i \wedge E_j, E_{i'} \wedge E_{j'}\right\rangle = \delta_{(i,j)(i',j')}\varepsilon_i\varepsilon_j.$$

In particular, we see that the inner product $\langle\cdot,\cdot\rangle$ on $\Lambda^2(V)$ is non-degenerate.

Definition 6.5. Let g, h be symmetric bilinear forms on V. We use g, h to define a $(0,4)$-tensor on V by setting

$$(g \boxslash h)(X, Y, W, Z) := \det \begin{pmatrix} g(X, W) & g(X, Z) \\ h(Y, W) & h(Y, Z) \end{pmatrix}.$$

Lemma 6.6. *$g \boxslash h$ satisfies the Bianchi identity.*

Proof. We compute

$$\begin{aligned}
(g \boxslash h)(X, Y, W, Z) &+ (g \boxslash h)(X, Z, Y, W) + (g \boxslash h)(X, W, Z, Y) \\
=& g(X, W)h(Y, Z) - g(X, Z)h(Y, W) \\
&+ g(X, Y)h(Z, W) - g(X, W)h(Z, Y) \\
&+ g(X, Z)h(Y, W) - g(X, Y)h(W, Z) \\
=& 0.
\end{aligned}$$

□

In general $g \boxslash h$ does not satisfy the skew symmetry from the definition of algebraic curvature tensors. However we see that for $g = h$ the skew symmetry holds and thus we use polarization:

$$(g + h) \boxslash (g + h) = g \boxslash g + g \boxslash h + h \boxslash g + h \boxslash h.$$

Since $(g + h) \boxslash (g + h)$, $g \boxslash g$ and $h \boxslash h$ are skew-symmetric we see that

$$g \boxslash h + h \boxslash g =: g \owedge h$$

is skew-symmetric and therefore defines an algebraic curvature tensor. The $(0,4)$-tensor $g \owedge h$ is called the *Kulkarni-Nomizu product* of g and h.

Lemma 6.7. *Let $g := \langle\cdot,\cdot\rangle$ be an inner product on V and let $n = \dim V$. Let R be an algebraic curvature tensor on V and define $\mathrm{Ric} := C(R)$ and $\mathrm{scal} := C(\mathrm{Ric}) := \langle \mathrm{Ric}, g\rangle$. Then the following holds*

(a) $C(g \boxslash g) = \frac{1}{2}C(g \owedge g) = (n-1)g$,
(b) $C(g \owedge \mathrm{Ric}) = (n-2)\mathrm{Ric} + \mathrm{scal} \cdot g$.

Proof. (a) We compute

$$\begin{aligned} C(g \boxslash g)(X,Y) &= \sum_{i=1}^{n} \varepsilon_i (g \boxslash g)(E_i, X, E_i, Y) \\ &= \sum_{i=1}^{n} \varepsilon_i [\langle E_i, E_i\rangle \langle X, Y\rangle - \langle E_i, Y\rangle \langle X, E_i\rangle] \\ &= n\langle X, Y\rangle - \langle X, Y\rangle \\ &= (n-1)g(X,Y). \end{aligned}$$

(b) We calculate

$$\begin{aligned} C(g \owedge \mathrm{Ric})(X,Y) &= \sum_{i=1}^{n} \varepsilon_i (g \boxslash \mathrm{Ric} + \mathrm{Ric} \boxslash g)(E_i, X, E_i, Y) \\ &= \sum_{i=1}^{n} \varepsilon_i [\langle E_i, E_i\rangle \mathrm{Ric}(X,Y) - \langle E_i, Y\rangle \mathrm{Ric}(X, E_i) \\ &\quad + \mathrm{Ric}(E_i, E_i)\langle X, Y\rangle - \mathrm{Ric}(E_i, Y)\langle X, E_i\rangle] \\ &= n\,\mathrm{Ric}(X,Y) - \mathrm{Ric}(X,Y) + \mathrm{scal}\,\langle X, Y\rangle - \mathrm{Ric}(X,Y) \\ &= (n-2)\mathrm{Ric}(X,Y) + \mathrm{scal}\cdot g(X,Y). \end{aligned}$$

□

Our next aim is to decompose every algebraic curvature tensor R as

$$R = W + \alpha g \owedge \mathrm{Ric} + \beta \mathrm{scal} \cdot g \owedge g,$$

where W is an algebraic curvature tensor with $C(W) = 0$ and $\alpha, \beta \in \mathbb{R}$ are universal constants which depend only on $n = \dim V$. Assume that we have such a decomposition. Then by the previous lemma we get

$$\begin{aligned} \mathrm{Ric} = C(R) &= C(W) + \alpha C(g \owedge \mathrm{Ric}) + \beta \mathrm{scal} \cdot C(g \owedge g) \\ &= \alpha((n-2)\mathrm{Ric} + \mathrm{scal} \cdot g) + 2\beta \mathrm{scal}(n-1)g \end{aligned}$$

and thus

$$(1 - \alpha(n-2))\mathrm{Ric} = (\alpha + 2\beta(n-1))\mathrm{scal} \cdot g.$$

If $n \geq 3$ then one can easily find examples of algebraic curvature tensors R such that Ric is not a multiple of g. Therefore we need to require $1 - \alpha(n-2) = 0$. It follows that $\alpha = \frac{1}{n-2}$ and $\frac{1}{n-2} + 2\beta(n-1) = 0$ thus $\beta = -\frac{1}{2(n-1)(n-2)}$.

Definition 6.8. For $n \geq 3$ the algebraic curvature tensor

$$W := R - \frac{1}{n-2} g \owedge \mathrm{Ric} + \frac{1}{2(n-1)(n-2)} \mathrm{scal} \cdot g \owedge g$$

is called the *Weyl curvature tensor* of R.

Corollary 6.9. *W is an algebraic curvature tensor with $C(W) = 0$.*

Remark 6.10. We can give an alternative decomposition of the algebraic curvature tensor R as follows. We want to decompose Ric as

$$\mathrm{Ric} = \mathrm{Ric}_0 + \gamma \mathrm{scal} \cdot g$$

such that $C(\mathrm{Ric}_0) = 0$ and $\gamma \in \mathbb{R}$. Assume that we have such a decomposition. It follows that

$$\mathrm{scal} = C(\mathrm{Ric}) = 0 + \gamma \mathrm{scal} \cdot n$$

and thus $\gamma = \frac{1}{n}$. Thus we define

$$\mathrm{Ric}_0 = \mathrm{Ric} - \frac{1}{n}\mathrm{scal} \cdot g.$$

Inserting this definition into the above decomposition of R yields

$$\begin{aligned} R &= W + \frac{1}{n-2} g \owedge \mathrm{Ric} - \frac{1}{2(n-1)(n-2)}\mathrm{scal} \cdot g \owedge g \\ &= W + \frac{1}{n-2} g \owedge \left(\mathrm{Ric}_0 + \frac{1}{n}\mathrm{scal} \cdot g\right) - \frac{1}{2(n-1)(n-2)}\mathrm{scal} \cdot g \owedge g \\ &= W + \frac{1}{n-2} g \owedge \mathrm{Ric}_0 + \left(\frac{1}{n(n-2)} - \frac{1}{2(n-1)(n-2)}\right)\mathrm{scal} \cdot g \owedge g \\ &= W + \frac{1}{n-2} g \owedge \mathrm{Ric}_0 + \frac{1}{2n(n-1)}\mathrm{scal} \cdot g \owedge g. \end{aligned}$$

Clearly, if $\mathrm{Ric} = 0$ then we have $R = W$.
If $n = 2$, every algebraic curvature tensor is of the form $R = \alpha g \owedge g$ with $\alpha \in \mathbb{R}$. Then we have $\mathrm{Ric} = 2\alpha g$, $\mathrm{Ric}_0 = 0$, $W = 0$ and $\mathrm{scal} = 4\alpha$, hence

$$R = \frac{\mathrm{scal}}{4} g \owedge g = \frac{K}{2} g \owedge g$$

where K is the Gauss curvature of the surface.

6.2. Hodge-⋆ operator

Let V be a real vector space of dimension n equipped with a non-degenerate inner product $\langle \cdot, \cdot \rangle$ of signature $(n - \sigma, \sigma)$ where $\sigma \in \{0, ..., n\}$ denotes the maximal dimension of a negative definite linear subspace of V. We assume that V is equipped with an orientation given by $\omega \in \Lambda^n(V)$ such that $\langle \omega, \omega \rangle = (-1)^\sigma$.

Lemma 6.11. *For every $k \in \{0, ..., n\}$ there is a unique homomorphism $\star : \Lambda^k(V) \to \Lambda^{n-k}(V)$ such that $\alpha \wedge \star\beta = \langle\alpha,\beta\rangle\,\omega$ for all $\alpha,\beta \in \Lambda^k(V)$.*

Proof. Uniqueness: Assume that $\star$ and $\tilde{\star}$ are two such homomorphisms. We fix $\beta \in \Lambda^k(V)$. Then for all $\alpha \in \Lambda^k(V)$ we have

$$\alpha \wedge \star\beta = \langle\alpha,\beta\rangle\,\omega = \alpha \wedge \tilde{\star}\beta$$

and thus $\alpha \wedge (\star\beta - \tilde{\star}\beta) = 0$. Let $(e_i)_{i=1}^n$ be an orthonormal basis of V. Then the elements of the form

$$e_{i_1} \wedge ... \wedge e_{i_k} =: e_I, \quad I = \{i_1, ..., i_k\}, \quad 1 \le i_1 < ... < i_k \le n,$$

are an orthonormal basis of $\Lambda^k(V)$. Thus we can write

$$\star\beta - \tilde{\star}\beta = \sum_I c_I e_I, \quad c_I \in \mathbb{R}, \quad I = \{i_1, ..., i_{n-k}\}, \quad 1 \le i_1 < ... < i_{n-k} \le n.$$

For every subset $J \subset \{1, ..., n\}$, $|J| = k$, we write $J^c := \{1, ..., n\} \setminus J$ and putting $\alpha := e_J$ we get

$$0 = \alpha \wedge (\star\beta - \tilde{\star}\beta) = e_J \wedge \sum_{I,\ |I|=n-k} c_I e_I = e_J \wedge c_{J^c} e_{J^c} = \pm c_{J^c}\omega.$$

Hence we have $c_{J^c} = 0$ for all J with $|J| = k$ and thus $\star\beta - \tilde{\star}\beta = 0$.
Existence: Let $I \subset \{1, ..., n\}$ with $|I| = k$. We define $\star e_I := \eta_I e_{I^c}$, where $\eta_I \in \{\pm 1\}$ is such that $\eta_I e_I \wedge e_{I^c} = \langle e_I, e_I\rangle\,\omega$. We extend this definition linearly and we obtain a map $\star\colon \Lambda^k(V) \to \Lambda^{n-k}(V)$ with the property stated in the assertion. □

Definition 6.12. The linear map $\star : \Lambda^k(V) \to \Lambda^{n-k}(V)$ is called the *Hodge-$\star$ operator.*

Since we will be interested in Lorentzian 4-manifolds we consider the following example.

Example 6.13. Let $n = 4$, $k = 2$, $\sigma = 1$. Let $e_1, ..., e_4$ be an orthonormal basis of V such that for $\varepsilon_i := \langle e_i, e_i\rangle$, $i = 1, ..., 4$ we have $\varepsilon_1 = -1$, $\varepsilon_2 = \varepsilon_3 = \varepsilon_4 = 1$ and let $\omega = -e_{1234}$. We compute

$$\begin{aligned}
e_{12} \wedge \star e_{12} &= \langle e_{12}, e_{12}\rangle\,\omega = (-1)(-e_{1234}) = e_{1234} \text{ and thus } \star e_{12} = e_{34},\\
e_{13} \wedge \star e_{13} &= \langle e_{13}, e_{13}\rangle\,\omega = e_{1234} = -e_{1324} \text{ and thus } \star e_{13} = -e_{24},\\
e_{14} \wedge \star e_{14} &= \langle e_{14}, e_{14}\rangle\,\omega = e_{1234} = e_{1423} \text{ and thus } \star e_{14} = e_{23},\\
e_{23} \wedge \star e_{23} &= \langle e_{23}, e_{23}\rangle\,\omega = -e_{1234} = -e_{2314} \text{ and thus } \star e_{23} = -e_{14}\\
e_{24} \wedge \star e_{24} &= \langle e_{24}, e_{24}\rangle\,\omega = -e_{1234} = e_{2413} \text{ and thus } \star e_{24} = e_{13},\\
e_{34} \wedge \star e_{34} &= \langle e_{34}, e_{34}\rangle\,\omega = -e_{1234} = -e_{3412} \text{ and thus } \star e_{34} = -e_{12}.
\end{aligned}$$

The general formula for $\{i, j, k, l\} = \{1, 2, 3, 4\}$ is

$$\star e_{ij} = -\varepsilon_i\varepsilon_j \operatorname{sign}(ijkl) e_{kl}\,.$$

Lemma 6.14. **(1)** *For all $\alpha \in \Lambda^k(V)$, $\beta \in \Lambda^{n-k}(V)$ we have $\langle \alpha, \star\beta\rangle = (-1)^{k(n-k)} \langle \star\alpha, \beta\rangle$.*
(2) *We have $\star^2 = (-1)^{k(n-k)+\sigma} \mathrm{id}_{\Lambda^k(V)}$.*

Proof. (1) Using the definition of $\star$ and the antisymmetry of $\wedge$ we get

$$\langle \star\alpha, \beta\rangle\, \omega = \star\alpha \wedge \star\beta = (-1)^{k(n-k)} \star\beta \wedge \star\alpha = (-1)^{k(n-k)} \langle \star\beta, \alpha\rangle\, \omega = (-1)^{k(n-k)} \langle \alpha, \star\beta\rangle\, \omega.$$

(2) For $I \subset \{1, ..., n\}$ with $|I| = k$ we have $\star e_I = \eta_I e_{I^c}$ with $\eta_I \in \{\pm 1\}$ and thus $\star \star e_I = \eta_I \eta_{I^c} e_I$. Taking the inner product with e_I we get

$$\eta_I \eta_{I^c} \langle e_I, e_I\rangle = \langle e_I, \star \star e_I\rangle = (-1)^{k(n-k)} \langle \star e_I, \star e_I\rangle = (-1)^{k(n-k)} \langle e_{I^c}, e_{I^c}\rangle$$

where in the last equality we have used $\star e_I = \eta_I e_{I^c}$ with $\eta_I \in \{\pm 1\}$. We get

$$\eta_I \eta_{I^c} = (-1)^{k(n-k)} \langle e_{I^c}, e_{I^c}\rangle \langle e_I, e_I\rangle$$

and by definition of σ we have $\langle e_{I^c}, e_{I^c}\rangle \langle e_I, e_I\rangle = (-1)^{\sigma}$. □

Example 6.15. Let $n = 4$, $k = 2$, $\sigma = 1$. Then $\star$ is self-adjoint with respect to $\langle \cdot, \cdot\rangle$ and $\star^2 = -\mathrm{id}_{\Lambda^2(V)}$. We define a complex structure on $\Lambda^2(V)$ by putting

$$(a + ib) \cdot \beta := a \cdot \beta + b \cdot \star\beta, \quad \beta \in \Lambda^2(V), \quad a, b \in \mathbb{R}.$$

Then $\Lambda^2(V)$ is a complex vector space with $\dim_{\mathbb{C}} \Lambda^2(V) = 3$. A $\mathbb{C}$-basis of $\Lambda^2(V)$ is given for example by (e_{12}, e_{13}, e_{14}). We define for $\alpha, \beta \in \Lambda^2(V)$

$$g_{\mathbb{C}}(\alpha, \beta) := \langle \alpha, \beta\rangle - i \langle \alpha, \star\beta\rangle\,.$$

Lemma 6.16. *$g_{\mathbb{C}}$ is a non-degenerate symmetric $\mathbb{C}$-bilinear form on $\Lambda^2(V)$.*

Proof. (a) $g_{\mathbb{C}}$ is symmetric: We have

$$g_{\mathbb{C}}(\beta, \alpha) = \langle \beta, \alpha\rangle - i \langle \beta, \star\alpha\rangle = \langle \alpha, \beta\rangle - i \langle \star\beta, \alpha\rangle = \langle \alpha, \beta\rangle - i \langle \alpha, \star\beta\rangle = g_{\mathbb{C}}(\alpha, \beta).$$

(b) $g_{\mathbb{C}}$ is $\mathbb{C}$-bilinear: It is clear that $g_{\mathbb{C}}$ is $\mathbb{R}$-bilinear. Because of the symmetry it is sufficient to prove complex linearity with respect to the second argument. Since $\star^2 = -\mathrm{id}_{\Lambda^2(V)}$ we have

$$g_{\mathbb{C}}(\alpha, i\beta) = \langle \alpha, \star\beta\rangle - i \langle \alpha, \star \star \beta\rangle = i(-i \langle \alpha, \star\beta\rangle + \langle \alpha, \beta\rangle) = i g_{\mathbb{C}}(\alpha, \beta).$$

(c) $g_{\mathbb{C}}$ is non-degenerate: If $g_{\mathbb{C}}(\alpha, \beta) = 0$ for all $\alpha \in \Lambda^2(V)$ then, taking real parts, we get $\langle \alpha, \beta\rangle = 0$ for all $\alpha \in \Lambda^2(V)$. Since $\langle \cdot, \cdot\rangle$ is non-degenerate we have $\beta = 0$. □

Thus $g_{\mathbb{C}}$ is a $\mathbb{C}$-bilinear extension of $\langle \cdot, \cdot\rangle$ on $\Lambda^2(V)$.

Remark 6.17. **(1)** An $\mathbb{R}$-linear map $A\colon \Lambda^2(V) \to \Lambda^2(V)$ is $\mathbb{C}$-linear if and only if we have $A \circ \star = \star \circ A$.
(2) If $A\colon \Lambda^2(V) \to \Lambda^2(V)$ is $\mathbb{C}$-linear and self-adjoint with respect to $\langle\cdot,\cdot\rangle$, then A is self-adjoint with respect to $g_{\mathbb{C}}$. Namely we have for all $\alpha,\beta \in \Lambda^2(V)$:

$$\begin{aligned} g_{\mathbb{C}}(A\alpha,\beta) &= \langle A\alpha,\beta\rangle - i\,\langle A\alpha,\star\beta\rangle = \langle\alpha,A\beta\rangle - i\,\langle\alpha,A\star\beta\rangle \\ &= \langle\alpha,A\beta\rangle - i\,\langle\alpha,\star A\beta\rangle = g_{\mathbb{C}}(\alpha,A\beta). \end{aligned}$$

Definition 6.18. An element $\beta \in \Lambda^k(V)$ is called *decomposable* if there exist $v_1,...,v_k \in V$ such that $\beta = v_1 \wedge ... \wedge v_k$.

Remark 6.19. If $\beta \in \Lambda^k(V) \setminus \{0\}$ is decomposable, $\beta = v_1 \wedge ... \wedge v_k$, then $v_1,...,v_k$ are linearly independent and they generate a k-dimensional linear subspace $\mathbb{R}\langle v_1,...,v_k\rangle$ of V. Moreover β determines an orientation on $\mathbb{R}\langle v_1,...,v_k\rangle$. We have one-to-one-correspondences

$$\begin{aligned} \mathbb{R}_+\cdot\beta &\longleftrightarrow \mathbb{R}\langle v_1,...,v_k\rangle \text{ with orientation,} \\ \mathbb{R}\cdot\beta &\longleftrightarrow \mathbb{R}\langle v_1,...,v_k\rangle. \end{aligned}$$

Proposition 6.20. *Let $\alpha \in \Lambda^2(V)$ where* $\dim V = n \geq 2$. *Then α is decomposable if and only if $\alpha \wedge \alpha = 0$.*

Proof. We prove the assertion by induction on n.
$n = 2$: The assertion holds since every $\alpha \in \Lambda^2(V)$ is decomposable and satisfies $\alpha \wedge \alpha = 0$.
Inductive step: Let $n \geq 3$.
"$\Longrightarrow$" If α is decomposable, i.e. $\alpha = v \wedge w$, where $v, w \in V$, then $\alpha \wedge \alpha = v \wedge w \wedge v \wedge w = 0$.
"$\Longleftarrow$" Let $\alpha \wedge \alpha = 0$. We choose a basis $b_1,...,b_n$ of V and we write $\alpha = b_1 \wedge v + \beta$ where $v \in W := \mathbb{R}\langle b_2,...,b_n\rangle$ and $\beta \in \Lambda^2(W)$. It follows that

$$0 = \alpha \wedge \alpha = (b_1 \wedge v + \beta) \wedge (b_1 \wedge v + \beta) = 2b_1 \wedge v \wedge \beta + \beta \wedge \beta.$$

Since $b_1 \wedge v \wedge \beta$ and $\beta \wedge \beta$ are linearly independent we have $b_1 \wedge v \wedge \beta = 0$ and $\beta \wedge \beta = 0$. By the inductive hypothesis β is decomposable, i.e. $\beta = w \wedge z$, where $w, z \in W$. It follows that

$$0 = b_1 \wedge v \wedge \beta = b_1 \wedge v \wedge w \wedge z.$$

Therefore the vectors b_1, v, w, z are linearly dependent. Since $v, w, z \in W$ and $b_1 \notin W$ we get that v, w, z are linearly dependent.
Case 1: $\beta = 0$: Then $\alpha = b_1 \wedge v$ is decomposable.
Case 2: $\beta \neq 0$, $v = 0$: Then $\alpha = \beta$ is decomposable.
Case 3: $\beta \neq 0$, $v \neq 0$: Since $\beta \neq 0$ the vectors w, z are linearly independent and we have $v \in \mathbb{R}\langle w, z\rangle$. Since $v \neq 0$ there exists a vector v' such that v, v' form a basis of $\mathbb{R}\langle w, z\rangle$. Thus there

is a constant $c \in \mathbb{R} \setminus \{0\}$ such that $v \wedge v' = cw \wedge z = c\beta$. It follows that

$$\alpha = b_1 \wedge v + \frac{1}{c} v \wedge v' = \left(b_1 - \frac{1}{c} v'\right) \wedge v$$

and hence α is decomposable. □

Corollary 6.21. *Every $\alpha \in \Lambda^2(\mathbb{R}^3)$ is decomposable.*

Proof. This follows from the lemma since $\alpha \wedge \alpha \in \Lambda^4(\mathbb{R}^3) = 0$. □

Example 6.22. The proof of the lemma gives an explicit way of constructing the decomposition of a decomposable element. For example we get

$$e_{12} + e_{23} + e_{13} = \tfrac{1}{2}(e_1 + 2e_2 + e_3) \wedge (e_1 + 4e_2 + 3e_3).$$

We also see that the corollary is not true for $n \geq 4$. Namely $\alpha = e_{12} + e_{34} \in \Lambda^2(\mathbb{R}^4)$ is not decomposable since

$$\alpha \wedge \alpha = (e_{12} + e_{34}) \wedge (e_{12} + e_{34}) = e_{1212} + e_{1234} + e_{3412} + e_{3434} = 2e_{1234} \neq 0.$$

Lemma 6.23. *Let $n = 4$, $k = 2$, $\sigma = 1$. Then $\star$ exchanges the causal types, i.e. if α is spacelike and not zero (timelike, lightlike) then $\star\alpha$ is timelike (spacelike and not zero, lightlike).*

Proof. We have $\langle \star\beta, \star\beta \rangle = \langle \star\star\beta, \beta \rangle = -\langle \beta, \beta \rangle$. □

Lemma 6.24. *Let $n = 4$, $k = 2$, $\sigma = 1$ and let $\beta \in \Lambda^2(V)$. Then the following statements are equivalent:*

(i) β is decomposable,
(ii) $\star\beta$ is decomposable,
(iii) $\beta \perp \star\beta$.

In particular by the one-to-one-correspondences in Remark 6.19 we can say that $\star$ is a map from the set of all 2-dimensional planes in V to itself.

Proof. (*i*) $\Longleftrightarrow$ (*iii*): We have

$$\beta \text{ decomposable } \Longleftrightarrow 0 = -\beta \wedge \beta = \beta \wedge \star\star\beta = \langle \beta, \star\beta \rangle\, \omega \Longleftrightarrow \langle \beta, \star\beta \rangle = 0.$$

$(i) \iff (ii)$: By the equivalence of (i) and (iii) we have

$$\beta \text{ decomposable } \iff \beta \perp \star\beta \iff \star\star\beta \perp \star\beta \iff \star\beta \text{ decomposable},$$

where we have applied the equivalence of (i) and (iii) to $\star\beta$. □

If we think of the Hodge-$\star$ operator as a map on the space of 2-dimensional subspaces of a 4-dimensional space, then Lemma 6.24 tells us that spacelike and timelike planes are mapped to their orthogonal complement. What happens to lightlike planes in V under the action of $\star$?

Lemma 6.25. *Let $n = 4$, $k = 2$, $\sigma = 1$ and let $\beta \in \Lambda^2(V) \setminus \{0\}$. Then β is lightlike and decomposable if and only if $\beta = \ell \wedge x$, where $\ell \in V$ is lightlike, $x \in V$ is spacelike and $x \perp \ell$.*

Proof. "$\Longleftarrow$" It remains to prove that β is lightlike. Since $\langle \ell, \ell\rangle = 0 = \langle \ell, x\rangle$ we get

$$\langle\beta,\beta\rangle = \langle \ell \wedge x, \ell \wedge x\rangle = \langle \ell,\ell\rangle\langle x,x\rangle - \langle \ell,x\rangle\langle x,\ell\rangle = 0.$$

"$\Longrightarrow$" Since β is decomposable we may write $\beta = v \wedge w$ where $v, w \in V$. Since $\beta \neq 0$ we know that v, w are linearly independent. By definition of $\langle\beta,\beta\rangle$ and since β is lightlike we have

$$0 = \langle\beta,\beta\rangle = \langle v,v\rangle\langle w,w\rangle - \langle v,w\rangle\langle w,v\rangle = \det\begin{pmatrix} \langle v,v\rangle & \langle w,v\rangle \\ \langle v,w\rangle & \langle w,w\rangle \end{pmatrix}.$$

Hence the restriction of $\langle\cdot,\cdot\rangle$ to $\mathbb{R}\cdot\langle v,w\rangle$ is degenerate. By a basic result of linear algebra on Lorentzian vector spaces there exist $\ell, x \in \mathbb{R}\cdot\langle v,w\rangle$ such that ℓ is lightlike, x is spacelike and $\ell \wedge x = v \wedge w = \beta$. □

Lemma 6.26. *Let $n = 4$, $k = 2$ and $\sigma = 1$ and let $\beta = \ell \wedge x \in \Lambda^2(V)$ where ℓ is lightlike, x is spacelike, $x \neq 0$, and $\ell \perp x$. Then $\star\beta = \ell \wedge y$ where y is spacelike, $|y| = |x|$ and $y \perp x$.*

Proof. Since x is spacelike and not zero the vector space $x^\perp$ has Lorentzian signature and its dimension is 3. We choose an orthonormal basis e_1, e_2, e_3 of $x^\perp$ such that e_1 is timelike, e_2, e_3 are spacelike and such that $\ell = a(e_1 + e_2)$ for some $a \in \mathbb{R}$. We put $e_4 := \frac{x}{|x|}$. Then $e_1, ..., e_4$ is an orthonormal basis of V. Using our results from Example 6.13 we get

$$\begin{aligned}\star\beta = \star(\ell \wedge x) &= \star(a(e_1 + e_2) \wedge |x|e_4) = a|x|(\star e_{14} + \star e_{24}) \\ &= a|x|(e_{23} + e_{13}) = a|x|(e_2 + e_1) \wedge e_3 = \ell \wedge |x|e_3\end{aligned}$$

and we finish the proof by putting $y := |x|e_3$. □

6.3. Curvature as a complex linear endomorphism

In this section V will be a real vector space of dimension $n = 4$ equipped with a non-degenerate inner product $\langle \cdot, \cdot \rangle$ with $\sigma = 1$. We have seen that in this case the Hodge star operator $\star$ on bivectors is self-adjoint and satisfies $\star^2 = -\mathrm{id}_{\Lambda^2(V)}$.

Lemma 6.27. *Let W be an algebraic curvature tensor on V with contraction $C(W) = 0$. If we consider W as a symmetric bilinear form on $\Lambda^2(V)$, then we have*

$$W(\star\omega, \star\eta) = -W(\omega, \eta)$$

for all $\omega, \eta \in \Lambda^2(V)$.

Proof. Let $(e_i)_{i=1}^4$ be an orthonormal basis of V. Then $(e_{ij})_{1 \le i < j \le 4}$ is an orthonormal basis of $\Lambda^2(V)$. It suffices to prove the claim for $\omega = e_{ij}$, $\eta = e_{k\ell}$, $i < j$, $k < \ell$.
Case 1: i, j, k, ℓ *are pairwise distinct.*
Then $\star\omega = \varepsilon\eta$, where $\varepsilon \in \{\pm 1\}$. Thus we have

$$W(\star\omega, \star\eta) = W(\varepsilon\eta, \varepsilon \star \star\omega) = -W(\eta, \omega) = -W(\omega, \eta).$$

Case 2: $\#(\{i, j\} \cap \{k, \ell\}) = 1$.
We may assume without loss of generality that $i = k = 1$, $j = 2$, $\ell = 3$, otherwise we change the numbering of the basis vectors e_i. We also may assume that the orientation is given by $-e_{1234}$, otherwise we replace e_4 by $-e_4$. Then we have $\omega = e_{12}$, $\eta = e_{13}$ and with $\varepsilon_i = \langle e_i, e_i \rangle$ we get by Example 6.13 that $\star\omega = -\varepsilon_1\varepsilon_2 e_{34}$ and $\star\eta = \varepsilon_1\varepsilon_3 e_{24}$. We conclude that

$$\begin{aligned} 0 = C(W)(e_2, e_3) &= \sum_{j=1}^{4} \varepsilon_j W(e_j, e_2, e_j, e_3) = \varepsilon_1 W(e_1, e_2, e_1, e_3) + \varepsilon_4 W(e_4, e_2, e_4, e_3) \\ &= \varepsilon_1 W(e_{12}, e_{13}) + \varepsilon_4 W(e_{24}, e_{34}). \end{aligned}$$

We multiply both sides by ε_1 and since $\varepsilon_1\varepsilon_2\varepsilon_3\varepsilon_4 = -1$ we get

$$\begin{aligned} 0 &= W(e_{12}, e_{13}) + \varepsilon_1\varepsilon_4 W(e_{24}, e_{34}) = W(\omega, \eta) + \varepsilon_1\varepsilon_4 W(\varepsilon_1\varepsilon_3 \star \eta, -\varepsilon_1\varepsilon_2 \star \omega) \\ &= W(\omega, \eta) - \varepsilon_1\varepsilon_2\varepsilon_3\varepsilon_4 W(\star\eta, \star\omega) = W(\omega, \eta) + W(\star\omega, \star\eta). \end{aligned}$$

Case 3: $i = k$, $j = \ell$.
We may assume without loss of generality that $i = k = 1$, $j = \ell = 2$. By hypothesis we have

$$0 = C(W)(e_1, e_1) = \varepsilon_2 W(e_{12}, e_{12}) + \varepsilon_3 W(e_{13}, e_{13}) + \varepsilon_4 W(e_{14}, e_{14}) \tag{1}$$
$$0 = C(W)(e_2, e_2) = \varepsilon_1 W(e_{12}, e_{12}) + \varepsilon_3 W(e_{23}, e_{23}) + \varepsilon_4 W(e_{24}, e_{24}) \tag{2}$$
$$0 = C(W)(e_3, e_3) = \varepsilon_1 W(e_{13}, e_{13}) + \varepsilon_2 W(e_{23}, e_{23}) + \varepsilon_4 W(e_{34}, e_{34}) \tag{3}$$
$$0 = C(W)(e_4, e_4) = \varepsilon_1 W(e_{14}, e_{14}) + \varepsilon_2 W(e_{24}, e_{24}) + \varepsilon_3 W(e_{34}, e_{34}). \tag{4}$$

We consider $\varepsilon_1 \cdot (1) + \varepsilon_2 \cdot (2) - \varepsilon_3 \cdot (3) - \varepsilon_4 \cdot (4)$ and we get

$$0 = 2\varepsilon_1\varepsilon_2 W(e_{12}, e_{12}) - 2\varepsilon_3\varepsilon_4 W(e_{34}, e_{34}).$$

We multiply both sides by $\frac{1}{2}\varepsilon_1\varepsilon_2$ and since $\varepsilon_1\varepsilon_2\varepsilon_3\varepsilon_4 = -1$ we get

$$0 = W(e_{12}, e_{12}) - \varepsilon_1\varepsilon_2\varepsilon_3\varepsilon_4 W(e_{34}, e_{34}) = W(\omega, \eta) + W(\star\omega, \star\eta)$$

which finishes the proof. □

Let $\mathcal{W}$ be the endomorphism of $\Lambda^2(V)$ corresponding to W, i.e. for all $\omega, \eta \in \Lambda^2(V)$ we have

$$\langle \mathcal{W}\omega, \eta \rangle = W(\omega, \eta).$$

Then $\mathcal{W}$ is self-adjoint with respect to $\langle \cdot, \cdot \rangle$. If we assume moreover that $C(W) = 0$, then by the previous lemma we get for all $\omega, \eta \in \Lambda^2(V)$

$$\langle \mathcal{W}(\star\omega), \eta \rangle = W(\star\omega, \eta) = -W(\star\star\omega, \star\eta) = W(\omega, \star\eta) = \langle \mathcal{W}\omega, \star\eta \rangle = \langle \star\mathcal{W}\omega, \eta \rangle .$$

Thus $\mathcal{W} \circ \star = \star \circ \mathcal{W}$, i.e. $\mathcal{W}$ is $\mathbb{C}$-linear.

Corollary 6.28. *Let $n = 4$, $\sigma = 1$. If W is an algebraic curvature tensor with $C(W) = 0$, then $\mathcal{W}$ is a $\mathbb{C}$-linear endomorphism of $\Lambda^2(V)$.*

We choose an orthonormal basis $(e_i)_{i=1}^4$ of V such that $\varepsilon_1 = -1$, $\varepsilon_2 = \varepsilon_3 = \varepsilon_4 = 1$. Then (e_{12}, e_{13}, e_{14}) form a complex basis of $\Lambda^2(V)$. We claim that the matrix of $\mathcal{W}$ with respect to this basis is complex and symmetric. In order to see this note first that $(e_{12}, e_{13}, e_{14}, ie_{12}, ie_{13}, ie_{14})$ form a real basis of $\Lambda^2(V)$ which is orthonormal with respect to $\langle \cdot, \cdot \rangle$ and that by definition of the complex structure on $\Lambda^2(V)$ we have $\star e_{1j} = ie_{1j}$, $j = 2, 3, 4$. Thus for $\omega \in \Lambda^2(V)$ we have by definition of $g_{\mathbb{C}}$

$$\omega = \sum_{j=2}^{4} \left(\left\langle \omega, e_{1j} \right\rangle e_{1j} + \left\langle \omega, \star e_{1j} \right\rangle ie_{1j} \right) = \sum_{j=2}^{4} \overline{g_{\mathbb{C}}(\omega, e_{1j})} e_{1j}.$$

Taking $\omega = \mathcal{W}e_{1k}$ we see that the matrix entries of $\mathcal{W}$ with respect to the basis $(e_{1j})_{j=2,3,4}$ are given by

$$\left(\overline{g_{\mathbb{C}}(\mathcal{W}e_{1k}, e_{1j})} \right)_{j,k=2,3,4}.$$

Now $\mathcal{W}$ is $\mathbb{C}$-linear and self-adjoint with respect to $\langle \cdot, \cdot \rangle$. By Remark 6.17 we conclude that $\mathcal{W}$ is self-adjoint with respect to $g_{\mathbb{C}}$. It follows that the matrix of $\mathcal{W}$ with respect to the basis (e_{12}, e_{13}, e_{14}) is complex and symmetric.

Lemma 6.29. *Let $n = 4$, $\sigma = 1$ and let W be an algebraic curvature tensor on V with $C(W) = 0$. Then we have $\mathrm{tr}_{\mathbb{C}}(\mathcal{W}) = 0$, where $\mathrm{tr}_{\mathbb{C}}(\mathcal{W})$ denotes the trace of $\mathcal{W}$ viewed as a complex linear endomorphism of $\Lambda^2(V)$.*

Proof. Let $(e_i)_{i=1}^4$ be an orthonormal basis of V such that $\varepsilon_1 = -1$, $\varepsilon_2 = \varepsilon_3 = \varepsilon_4 = 1$. We have $\mathrm{tr}_{\mathbb{C}}(\mathcal{W}) = \sum_{j=2}^4 \overline{g_{\mathbb{C}}(\mathcal{W}e_{1j}, e_{1j})}$ and thus by definition of $g_{\mathbb{C}}$

$$\mathrm{Re}(\mathrm{tr}_{\mathbb{C}}(\mathcal{W})) = \sum_{j=2}^4 \left\langle \mathcal{W}e_{1j}, e_{1j} \right\rangle = \sum_{j=2}^4 W(e_{1j}, e_{1j}) = C(W)(e_1, e_1) = 0$$

and

$$\begin{aligned} -\mathrm{Im}(\mathrm{tr}_{\mathbb{C}}(\mathcal{W})) &= \sum_{j=2}^4 \left\langle \mathcal{W}e_{1j}, \star e_{1j} \right\rangle = \sum_{j=2}^4 W(e_{1j}, \star e_{1j}) = W(e_{12}, e_{34}) + W(e_{13}, e_{42}) + W(e_{14}, e_{23}) \\ &= W(e_1, e_2, e_3, e_4) + W(e_1, e_3, e_4, e_2) + W(e_1, e_4, e_2, e_3) = 0 \end{aligned}$$

by the Bianchi identity. $\square$

We remark that complex symmetric matrices are not necessarily diagonalizable. For example the matrix

$$A = \begin{pmatrix} 1 & i \\ i & -1 \end{pmatrix}$$

is symmetric. Its characteristic polynomial is $\chi_A(\lambda) = (1-\lambda)(-1-\lambda) - i^2 = \lambda^2$ and thus 0 is the only eigenvalue of A. However A is not diagonalizable since otherwise we would have $A = 0$.

6.4. The Petrov types

The idea of Petrov classification is to divide the trace-free $\mathbb{C}$-linear endomorphisms of a 3-dimensional complex vector space into types according to the shape of their Jordan normal form. More precisely, let $\mathcal{W}$ be a $\mathbb{C}$-linear endomorphism of a 3-dimensional complex vector space whose trace is 0. We call $\mathcal{W}$ of type $X \in \{I, II, III, D, N, O\}$ if its Jordan normal form has the following shape:
If $\mathcal{W}$ is diagonalizable, i.e. $\mathcal{W}$ has 3 Jordan blocks

$$\begin{pmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \lambda_3 \end{pmatrix} \qquad \begin{pmatrix} \lambda_1 & & \\ & \lambda_1 & \\ & & \lambda_2 \end{pmatrix} \qquad \begin{pmatrix} 0 & & \\ & 0 & \\ & & 0 \end{pmatrix}$$

$\lambda_1, \lambda_2, \lambda_3$ pairwise distinct, $\lambda_1 + \lambda_2 + \lambda_3 = 0$ — type I

$\lambda_1 \neq \lambda_2 = -2\lambda_1$ — type D

$\lambda_1 = \lambda_2 = \lambda_3 = 0$ — type O

If $\mathcal{W}$ has 2 or 1 Jordan block

$$\begin{pmatrix} \lambda_1 & 1 & \\ & \lambda_1 & \\ & & \lambda_2 \end{pmatrix} \qquad \begin{pmatrix} 0 & 1 & \\ & 0 & \\ & & 0 \end{pmatrix} \qquad \begin{pmatrix} 0 & 1 & \\ & 0 & 1 \\ & & 0 \end{pmatrix}$$

$\lambda_1 \neq \lambda_2 = -2\lambda_1$ — type II

$\lambda_1 = \lambda_2 = \lambda_3 = 0$ — type N

$\lambda_1 = \lambda_2 = \lambda_3 = 0$ — type III

In the following table we list some more properties of these types.

type	I	II	III	D	N	O
#(Jordan blocks)	3	2	1	3	2	3
diagonalizable	yes	no	no	yes	no	yes
distinct eigenvalues	$\lambda_1, \lambda_2, \lambda_3$	λ_1, λ_2	0	λ_1, λ_2	0	0
algebraic multiplicities	$\mu_{\text{alg}}(\lambda_j) = 1$	$\mu_{\text{alg}}(\lambda_1) = 2$ $\mu_{\text{alg}}(\lambda_2) = 1$	$\mu_{\text{alg}}(0) = 3$	$\mu_{\text{alg}}(\lambda_1) = 2$ $\mu_{\text{alg}}(\lambda_2) = 1$	$\mu_{\text{alg}}(0) = 3$	$\mu_{\text{alg}}(0) = 3$
geometric multiplicities	$\mu_{\text{geo}}(\lambda_j) = 1$	$\mu_{\text{geo}}(\lambda_1) = 1$ $\mu_{\text{geo}}(\lambda_2) = 1$	$\mu_{\text{geo}}(0) = 1$	$\mu_{\text{geo}}(\lambda_1) = 2$ $\mu_{\text{geo}}(\lambda_2) = 1$	$\mu_{\text{geo}}(0) = 2$	$\mu_{\text{geo}}(0) = 3$
rank	2 or 3	3	2	3	1	0
invertible	depends	yes	no	yes	no	no
nilpotent	no	no	yes	no	yes	yes
$\min\{k \mid A^k = 0\}$	-	-	3	-	2	1

Let $X, Y \in \{I, II, III, D, N, O\}$. We say that type X degenerates to type Y if there is a sequence of endomorphisms $(\mathcal{W}_n)_{n\in\mathbb{N}}$ all of type X and an endomorphism $\mathcal{W}$ of type Y such that $\mathcal{W}_n \to \mathcal{W}$ with respect to some (hence every) matrix norm on $\mathbb{C}^{3\times 3}$. In this case we write $X \to Y$. We claim that the following degenerations are possible:

$$\begin{tikzcd} I \arrow[d] \arrow[dr] & \\ II \arrow[r] \arrow[d] & D \arrow[d] \arrow[dr] & \\ III \arrow[r] & N \arrow[r] & O \end{tikzcd} \tag{5}$$

This follows by considering (as $\varepsilon \to 0$)

$$I \to D: \quad \begin{pmatrix} \lambda_1+\varepsilon & & \\ & \lambda_1-\varepsilon & \\ & & -2\lambda_1 \end{pmatrix} \to \begin{pmatrix} \lambda_1 & & \\ & \lambda_1 & \\ & & -2\lambda_1 \end{pmatrix}$$

$$I \to II: \quad \begin{pmatrix} \lambda_1 & 1 & \\ \varepsilon & \lambda_1 & \\ & & -2\lambda_1 \end{pmatrix} \to \begin{pmatrix} \lambda_1 & 1 & \\ & \lambda_1 & \\ & & -2\lambda_1 \end{pmatrix}$$

$$II \to D: \quad \begin{pmatrix} \lambda_1 & \varepsilon & \\ & \lambda_1 & \\ & & -2\lambda_1 \end{pmatrix} \to \begin{pmatrix} \lambda_1 & & \\ & \lambda_1 & \\ & & -2\lambda_1 \end{pmatrix}$$

$$II \to III: \quad \begin{pmatrix} \varepsilon & 1 & \\ & \varepsilon & 1 \\ & & -2\varepsilon \end{pmatrix} \to \begin{pmatrix} 0 & 1 & \\ & 0 & 1 \\ & & 0 \end{pmatrix}$$

$$D \to N: \quad \begin{pmatrix} 2\varepsilon & 1 & \\ & -\varepsilon & \\ & & -\varepsilon \end{pmatrix} \to \begin{pmatrix} 0 & 1 & \\ & 0 & \\ & & 0 \end{pmatrix}$$

$$D \to O: \quad \begin{pmatrix} \varepsilon & & \\ & \varepsilon & \\ & & -2\varepsilon \end{pmatrix} \to \begin{pmatrix} 0 & & \\ & 0 & \\ & & 0 \end{pmatrix}$$

$$III \to N: \quad \begin{pmatrix} 0 & 1 & \\ & 0 & \varepsilon \\ & & 0 \end{pmatrix} \to \begin{pmatrix} 0 & 1 & \\ & 0 & \\ & & 0 \end{pmatrix}$$

$$N \to O: \quad \begin{pmatrix} 0 & \varepsilon & \\ & 0 & \\ & & 0 \end{pmatrix} \to \begin{pmatrix} 0 & & \\ & 0 & \\ & & 0 \end{pmatrix}$$

One can show that type X degenerates to type Y if and only if Y can be reached from X by a sequence of arrows in diagram (5).
We say that an algebraic curvature tensor W with $C(W) = 0$ has *Petrov type* $X \in \{I, II, III, D, N, O\}$ if the corresponding $\mathbb{C}$-linear endomorphism $\mathcal{W}$ has type X.

Remark 6.30. The Petrov type of an algebraic curvature tensor W with $C(W) = 0$ is independent of the choice of orientation on V. Namely, if we reverse the orientation on V, then $\star$ is replaced by $-\star$, thus $g_{\mathbb{C}}$ is replaced by $\overline{g_{\mathbb{C}}}$. Therefore the matrix of $\mathcal{W}$ with respect to a basis as above is replaced by its complex conjugate matrix. It follows that the type of W does not change.

Definition 6.31. Let M be a 4-dimensional Lorentzian manifold. We say that M has *Petrov type* $X \in \{I, II, III, D, N, O\}$ *at the point* $p \in M$ if the Weyl tensor $W(p)$ has type X. If M has the same type X at all its points we say that *M has type X*.

Example 6.32. For $m > 0$ we consider $M = \mathbb{R} \times ((0, 2m) \cup (2m, \infty)) \times S^2$ equipped with the Schwarzschild metric g_m which in coordinates (t, r) on $\mathbb{R} \times ((0, 2m) \cup (2m, \infty)$ and spherical coordinates (ϑ, φ) on $S^2 \setminus \{0, 0, \pm 1\}$ is given by

$$g_m = -h(r)dt^2 + \frac{1}{h(r)}dr^2 + r^2 d\vartheta^2 + r^2 \sin^2\vartheta d\varphi^2, \quad h(r) := 1 - \frac{2m}{r}.$$

We know that $\mathrm{Ric} = 0$ and thus the Riemann curvature tensor R is equal to the Weyl curvature tensor W everywhere. For every point $p \in M$ whose S^2-component is different from $\{0, 0, \pm 1\}$ we consider the following orthonormal basis $(E_i)_{i=1}^4$ of T_pM with $\varepsilon_i = \langle E_i, E_i \rangle$:

	$r > 2m$	$r < 2m$
$E_1 := \lvert h(r)\rvert^{-1/2}\partial_t$	$\varepsilon_1 = -1$	$\varepsilon_1 = 1$
$E_2 := \lvert h(r)\rvert^{1/2}\partial_r$	$\varepsilon_2 = 1$	$\varepsilon_2 = -1$
$E_3 := \frac{1}{r}\partial_\vartheta$	$\varepsilon_3 = 1$	$\varepsilon_3 = 1$
$E_4 := \frac{1}{r\sin\vartheta}\partial_\varphi$	$\varepsilon_4 = 1$	$\varepsilon_4 = 1$

We denote by $(e_i)_{i=1}^4$ the basis of T_p^*M dual to this basis. One can show that the endomorphism $\mathcal{W}$ of $\Lambda^2 T_p^*M$ which corresponds to $W(p)$ has the eigenvalues

$$\lambda_1 = -\frac{m}{r^3}, \quad \lambda_2 = \frac{2m}{r^3}$$

and that the eigenspaces are

$$\begin{aligned}\mathrm{Eig}(\mathcal{W}, \lambda_1) &= \mathrm{span}_{\mathbb{R}}\{dt \wedge d\vartheta, dt \wedge d\varphi, dr \wedge d\vartheta, dr \wedge d\varphi\} = \mathrm{span}_{\mathbb{R}}\{e_{13}, e_{14}, e_{23}, e_{24}\},\\ \mathrm{Eig}(\mathcal{W}, \lambda_2) &= \mathrm{span}_{\mathbb{R}}\{dt \wedge dr, d\vartheta \wedge d\varphi\} = \mathrm{span}_{\mathbb{R}}\{e_{12}, e_{34}\}.\end{aligned}$$

If we consider $\Lambda^2 T_p^*M$ as a complex vector space then the endomorphism $\mathcal{W}$ is $\mathbb{C}$-linear and its matrix with respect to a basis of eigenvectors is given by

$r > 2m$

with respect to the basis e_{12}, e_{13}, e_{14}:

$$\begin{pmatrix} \frac{2m}{r^3} & & \\ & -\frac{m}{r^3} & \\ & & -\frac{m}{r^3} \end{pmatrix}$$

$r < 2m$

with respect to the basis e_{21}, e_{23}, e_{24}:

$$\begin{pmatrix} \frac{2m}{r^3} & & \\ & -\frac{m}{r^3} & \\ & & -\frac{m}{r^3} \end{pmatrix}$$

In particular, the Schwarzschild metric g_m has Petrov type D. We recall that every family of Kerr metrics $(g_{(m,a)})_{a>0}$ tends to the Schwarzschild metric g_m as $a \to 0$. From the above result on possible degenerations it follows that every Kerr metric has type I, II or D for small a. In fact one can show that every Kerr metric has type D and one can compute the eigenvalues of $\mathcal{W}$ explicitly (see e.g. Corollary 5.4.4 in [8]).

Lemma 6.33. *Let $\widetilde{V}$ be a 3-dimensional complex vector space equipped with a non-degenerate symmetric $\mathbb{C}$-bilinear form $g_{\mathbb{C}}$. Let $\mathcal{W}$ be a $\mathbb{C}$-linear endomorphism of $\widetilde{V}$ which is symmetric with respect to $g_{\mathbb{C}}$ and satisfies* $\mathrm{tr}_{\mathbb{C}}(\mathcal{W}) = 0$. *Then there exists a basis $\omega_1, \omega_2, \omega_3$ of $\widetilde{V}$ with respect to which we have*

type	*matrix of $g_{\mathbb{C}}$*	*matrix of $\mathcal{W}$*
I, D, O	$\begin{pmatrix} 1 & & \\ & 1 & \\ & & 1 \end{pmatrix}$	$\begin{pmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \lambda_3 \end{pmatrix}$
II, N	$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$	$\begin{pmatrix} \lambda_1 & & \\ & \lambda_2 & 1 \\ & & \lambda_2 \end{pmatrix}$
III	$\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 1 & \\ & 0 & 1 \\ & & 0 \end{pmatrix}$

Proof. a) We first prove that eigenspaces corresponding to different eigenvalues are $g_{\mathbb{C}}$-orthogonal. Namely let $\omega, \eta \in \widetilde{V}$ be eigenvectors, $\mathcal{W}\omega = \lambda\omega$, $\mathcal{W}\eta = \mu\eta$, $\lambda \neq \mu$. It follows that

$$\lambda g_{\mathbb{C}}(\omega, \eta) = g_{\mathbb{C}}(\lambda\omega, \eta) = g_{\mathbb{C}}(\mathcal{W}\omega, \eta) = g_{\mathbb{C}}(\omega, \mathcal{W}\eta) = \mu g_{\mathbb{C}}(\omega, \eta).$$

Thus we get $(\lambda - \mu) g_{\mathbb{C}}(\omega, \eta) = 0$ and since $\lambda \neq \mu$ we obtain $g_{\mathbb{C}}(\omega, \eta) = 0$.
b) We claim: If $\mathcal{W}$ is diagonalizable then the restriction of $g_{\mathbb{C}}$ to every eigenspace of $\mathcal{W}$ is non-degenerate.
Namely $\widetilde{V}$ is the direct sum of the eigenspaces of $\mathcal{W}$ since $\mathcal{W}$ is diagonalizable. Let λ be an eigenvalue of $\mathcal{W}$. Then we have

$$\dim \bigoplus_{\mu \neq \lambda} \mathrm{Eig}(\mathcal{W}, \mu) = \dim \widetilde{V} - \dim \mathrm{Eig}(\mathcal{W}, \lambda) = \dim \mathrm{Eig}(\mathcal{W}, \lambda)^{\perp}.$$

By part a) we have

$$\bigoplus_{\mu \neq \lambda} \mathrm{Eig}(\mathcal{W}, \mu) \subset \mathrm{Eig}(\mathcal{W}, \lambda)^{\perp}$$

and since the dimensions agree the two spaces are equal. Assume that the restriction of $g_{\mathbb{C}}$ to $\mathrm{Eig}(\mathcal{W}, \lambda)$ is degenerate. Then we have $\mathrm{Eig}(\mathcal{W}, \lambda)^{\perp} \cap \mathrm{Eig}(\mathcal{W}, \lambda) \neq \{0\}$ and thus there exists

$$v \in \mathrm{Eig}(\mathcal{W}, \lambda) \cap \bigoplus_{\mu \neq \lambda} \mathrm{Eig}(\mathcal{W}, \mu), \quad v \neq 0,$$

which is impossible. This concludes the proof of our claim.
c) Consider types I, D, O. Then $\mathcal{W}$ is diagonalizable. For every eigenspace of $\mathcal{W}$ we can find an orthonormal basis by part b). Note that for every vector v of this basis we may assume that $g_{\mathbb{C}}(v, v) = +1$ since we can multiply v by a complex number. The collection of all these basis vectors gives an orthonormal basis of $\widetilde{V}$ with respect to which the matrix of $\mathcal{W}$ has the desired form.
d) Consider type II, i.e. $\lambda_1 \neq \lambda_2$. Let $\widetilde{\omega}_1, \widetilde{\omega}_2, \widetilde{\omega}_3$ be a basis of $\widetilde{V}$ such that the matrix of $\mathcal{W}$ with respect to this basis has the form stated above. In particular $\widetilde{\omega}_1$ is an eigenvector of $\mathcal{W}$ corresponding to λ_1 and $\widetilde{\omega}_2$ is an eigenvector of $\mathcal{W}$ corresponding to λ_2. Thus by part a) we have $g_{\mathbb{C}}(\widetilde{\omega}_1, \widetilde{\omega}_2) = 0$. Furthermore we compute

$$\begin{aligned} \lambda_1 g_{\mathbb{C}}(\widetilde{\omega}_1, \widetilde{\omega}_3) &= g_{\mathbb{C}}(\mathcal{W}\widetilde{\omega}_1, \widetilde{\omega}_3) = g_{\mathbb{C}}(\widetilde{\omega}_1, \mathcal{W}\widetilde{\omega}_3) = g_{\mathbb{C}}(\widetilde{\omega}_1, \widetilde{\omega}_2 + \lambda_2\widetilde{\omega}_3) \\ &= g_{\mathbb{C}}(\widetilde{\omega}_1, \widetilde{\omega}_2) + \lambda_2 g_{\mathbb{C}}(\widetilde{\omega}_1, \widetilde{\omega}_3) = \lambda_2 g_{\mathbb{C}}(\widetilde{\omega}_1, \widetilde{\omega}_3). \end{aligned}$$

Since $\lambda_1 \neq \lambda_2$ we get $g_{\mathbb{C}}(\widetilde{\omega}_1, \widetilde{\omega}_3) = 0$. We conclude that $g_{\mathbb{C}}(\widetilde{\omega}_1, \widetilde{\omega}_1) \neq 0$ since $g_{\mathbb{C}}$ is non-degenerate. We define $\omega_1 := a\widetilde{\omega}_1$ where $a \in \mathbb{C}$ is such that $g_{\mathbb{C}}(\omega_1, \omega_1) = a^2 g_{\mathbb{C}}(\widetilde{\omega}_1, \widetilde{\omega}_1) = 1$. Next we compute

$$\begin{aligned} \lambda_2 g_{\mathbb{C}}(\widetilde{\omega}_2, \widetilde{\omega}_3) &= g_{\mathbb{C}}(\mathcal{W}\widetilde{\omega}_2, \widetilde{\omega}_3) = g_{\mathbb{C}}(\widetilde{\omega}_2, \mathcal{W}\widetilde{\omega}_3) = g_{\mathbb{C}}(\widetilde{\omega}_2, \widetilde{\omega}_2 + \lambda_2\widetilde{\omega}_3) \\ &= g_{\mathbb{C}}(\widetilde{\omega}_2, \widetilde{\omega}_2) + \lambda_2 g_{\mathbb{C}}(\widetilde{\omega}_2, \widetilde{\omega}_3). \end{aligned}$$

It follows that $g_{\mathbb{C}}(\widetilde{\omega}_2, \widetilde{\omega}_2) = 0$ and since $g_{\mathbb{C}}$ is non-degenerate we conclude that $g_{\mathbb{C}}(\widetilde{\omega}_2, \widetilde{\omega}_3) \neq 0$. We define $\omega_2 := b\widetilde{\omega}_2$ and $\widehat{\omega}_3 := b\widetilde{\omega}_3$ where $b \in \mathbb{C}$ is such that $g_{\mathbb{C}}(\omega_2, \widehat{\omega}_3) = b^2 g_{\mathbb{C}}(\widetilde{\omega}_2, \widetilde{\omega}_3) = 1$. Note that we have $\mathcal{W}\widehat{\omega}_3 = \omega_2 + \lambda_2\widehat{\omega}_3$, i.e. the matrix of $\mathcal{W}$ with respect to the basis $\omega_1, \omega_2, \widehat{\omega}_3$ is still as claimed above. Finally we put $\omega_3 := \widehat{\omega}_3 + c\omega_2$ where $c \in \mathbb{C}$ is such that

$$g_{\mathbb{C}}(\omega_3, \omega_3) = g_{\mathbb{C}}(\widehat{\omega}_3, \widehat{\omega}_3) + 2c g_{\mathbb{C}}(\widehat{\omega}_3, \omega_2) + c^2 g_{\mathbb{C}}(\omega_2, \omega_2) = g_{\mathbb{C}}(\widehat{\omega}_3, \widehat{\omega}_3) + 2c \stackrel{!}{=} 1.$$

Then the matrix of $g_{\mathbb{C}}$ with respect to $\omega_1, \omega_2, \omega_3$ has the desired form. Note that we have $\mathcal{W}\omega_3 = \omega_2 + \lambda_2\widehat{\omega}_3 + c\lambda_2\omega_2 = \omega_2 + \lambda_2\omega_3$, i.e. the matrix of $\mathcal{W}$ with respect to $\omega_1, \omega_2, \omega_3$ is the same as before.

d) Consider type N, i.e. $\lambda_1 = \lambda_2 = 0$. Let $\widetilde{\omega}_1, \widetilde{\omega}_2, \widetilde{\omega}_3$ be a basis of $\widetilde{V}$ such that the matrix of $\mathcal{W}$ with respect to this basis is as claimed above. Since $\mathcal{W}\widetilde{\omega}_1 = \mathcal{W}\widetilde{\omega}_2 = 0$ we get

$$\begin{aligned} 0 &= g_{\mathbb{C}}(\mathcal{W}\widetilde{\omega}_1, \widetilde{\omega}_3) = g_{\mathbb{C}}(\widetilde{\omega}_1, \mathcal{W}\widetilde{\omega}_3) = g_{\mathbb{C}}(\widetilde{\omega}_1, \widetilde{\omega}_2), \\ 0 &= g_{\mathbb{C}}(\mathcal{W}\widetilde{\omega}_2, \widetilde{\omega}_3) = g_{\mathbb{C}}(\widetilde{\omega}_2, \mathcal{W}\widetilde{\omega}_3) = g_{\mathbb{C}}(\widetilde{\omega}_2, \widetilde{\omega}_2) \end{aligned}$$

and since $g_{\mathbb{C}}$ is non-degenerate we conclude that $g_{\mathbb{C}}(\widetilde{\omega}_2, \widetilde{\omega}_3) \neq 0$. After multiplying both $\widetilde{\omega}_2, \widetilde{\omega}_3$ by a common factor we may assume that $g_{\mathbb{C}}(\widetilde{\omega}_2, \widetilde{\omega}_3) = 1$ and we still have $\mathcal{W}\widetilde{\omega}_3 = \widetilde{\omega}_2$. If we had $g_{\mathbb{C}}(\widetilde{\omega}_1, \widetilde{\omega}_1) = 0$, then the restriction of $g_{\mathbb{C}}$ to $\ker \mathcal{W} \times \ker \mathcal{W}$ would be identically zero. Then we could find a linear combination of $\widetilde{\omega}_1, \widetilde{\omega}_2$ which is $g_{\mathbb{C}}$-orthogonal to $\widetilde{\omega}_3$ which is impossible since $g_{\mathbb{C}}$ is non-degenerate. Thus we have $g_{\mathbb{C}}(\widetilde{\omega}_1, \widetilde{\omega}_1) \neq 0$ and after multiplying $\widetilde{\omega}_1$ by a complex number we may assume that $g_{\mathbb{C}}(\widetilde{\omega}_1, \widetilde{\omega}_1) = 1$. Next we put

$$\omega_2 := \widetilde{\omega}_2, \quad \omega_3 := \widetilde{\omega}_3 + a\widetilde{\omega}_2, \quad \omega_1 := \widetilde{\omega}_1 + b\widetilde{\omega}_2$$

where we choose $a, b \in \mathbb{C}$ such that $g_{\mathbb{C}}(\omega_1, \omega_3) = g_{\mathbb{C}}(\omega_3, \omega_3) = 0$. This is possible since we have

$$0 \stackrel{!}{=} g_{\mathbb{C}}(\omega_3, \omega_3) = g_{\mathbb{C}}(\widetilde{\omega}_3, \widetilde{\omega}_3) + 2a g_{\mathbb{C}}(\widetilde{\omega}_3, \widetilde{\omega}_2) + a^2 g_{\mathbb{C}}(\widetilde{\omega}_2, \widetilde{\omega}_2) = g_{\mathbb{C}}(\widetilde{\omega}_3, \widetilde{\omega}_3) + 2a$$

and

$$\begin{aligned} 0 &\stackrel{!}{=} g_{\mathbb{C}}(\omega_1, \omega_3) = g_{\mathbb{C}}(\widetilde{\omega}_1 + b\widetilde{\omega}_2, \widetilde{\omega}_3 + a\widetilde{\omega}_2) \\ &= g_{\mathbb{C}}(\widetilde{\omega}_1, \widetilde{\omega}_3) + a g_{\mathbb{C}}(\widetilde{\omega}_1, \widetilde{\omega}_2) + b g_{\mathbb{C}}(\widetilde{\omega}_2, \widetilde{\omega}_3) + ab g_{\mathbb{C}}(\widetilde{\omega}_2, \widetilde{\omega}_2) = g_{\mathbb{C}}(\widetilde{\omega}_1, \widetilde{\omega}_3) + b. \end{aligned}$$

We note that $\mathcal{W}\omega_3 = \mathcal{W}\widetilde{\omega}_3 + a\mathcal{W}\widetilde{\omega}_2 = \widetilde{\omega}_2 = \omega_2$ and therefore the matrix of $\mathcal{W}$ with respect to the new basis is the same as before. We also note that

$$g_{\mathbb{C}}(\omega_1, \omega_1) = 1, \quad g_{\mathbb{C}}(\omega_2, \omega_3) = 1, \quad g_{\mathbb{C}}(\omega_1, \omega_2) = g_{\mathbb{C}}(\omega_2, \omega_2) = 0$$

and thus the matrix of $g_{\mathbb{C}}$ with respect to the new basis has the desired form.

e) Consider type III. Let $\widetilde{\omega}_1, \widetilde{\omega}_2, \widetilde{\omega}_3$ be a basis of $\widetilde{V}$ such that the matrix of $\mathcal{W}$ with respect to this basis is as claimed above. Note that we have $\mathcal{W}\widetilde{\omega}_3 = \omega_2$, $\mathcal{W}\widetilde{\omega}_2 = \omega_1$, $\mathcal{W}\widetilde{\omega}_1 = 0$ and $\mathcal{W}^3 = 0$. It follows that

$$\begin{aligned} g_{\mathbb{C}}(\widetilde{\omega}_1, \widetilde{\omega}_1) &= g_{\mathbb{C}}(\mathcal{W}^2\widetilde{\omega}_3, \mathcal{W}^2\widetilde{\omega}_3) = g_{\mathbb{C}}(\mathcal{W}^4\widetilde{\omega}_3, \widetilde{\omega}_3) = 0 \\ g_{\mathbb{C}}(\widetilde{\omega}_1, \widetilde{\omega}_2) &= g_{\mathbb{C}}(\mathcal{W}^2\widetilde{\omega}_3, \mathcal{W}\widetilde{\omega}_3) = g_{\mathbb{C}}(\mathcal{W}^3\widetilde{\omega}_3, \widetilde{\omega}_3) = 0. \end{aligned}$$

Therefore since $g_{\mathbb{C}}$ is non-degenerate we have $g_{\mathbb{C}}(\widetilde{\omega}_1,\widetilde{\omega}_3) \neq 0$. After multiplying the vectors $\widetilde{\omega}_1,\widetilde{\omega}_2,\widetilde{\omega}_3$ by the same factor we may assume that $g_{\mathbb{C}}(\widetilde{\omega}_1,\widetilde{\omega}_3) = 1$. Note that with respect to the new basis the matrix of $\mathcal{W}$ is the same as before. It follows that

$$g_{\mathbb{C}}(\widetilde{\omega}_2,\widetilde{\omega}_2) = g_{\mathbb{C}}(\mathcal{W}\widetilde{\omega}_3,\mathcal{W}\widetilde{\omega}_3) = g_{\mathbb{C}}(\mathcal{W}^2\widetilde{\omega}_3,\widetilde{\omega}_3) = g_{\mathbb{C}}(\widetilde{\omega}_1,\widetilde{\omega}_3) = 1.$$

Next we define

$$\omega_1 := \widetilde{\omega}_1, \quad \omega_2 := \widetilde{\omega}_2 + a\widetilde{\omega}_1, \quad \omega_3 := \widetilde{\omega}_3 + a\widetilde{\omega}_2 + b\widetilde{\omega}_1$$

where we choose $a, b \in \mathbb{C}$ such that $g_{\mathbb{C}}(\omega_2,\omega_3) = 0 = g_{\mathbb{C}}(\omega_3,\omega_3)$. This is possible since we have

$$\begin{aligned}
0 &\overset{!}{=} g_{\mathbb{C}}(\omega_2,\omega_3) = g_{\mathbb{C}}(\widetilde{\omega}_2 + a\widetilde{\omega}_1, \widetilde{\omega}_3 + a\widetilde{\omega}_2 + b\widetilde{\omega}_1) \\
&= g_{\mathbb{C}}(\widetilde{\omega}_2,\widetilde{\omega}_3) + ag_{\mathbb{C}}(\widetilde{\omega}_1,\widetilde{\omega}_3) + ag_{\mathbb{C}}(\widetilde{\omega}_2,\widetilde{\omega}_2) + a^2 g_{\mathbb{C}}(\widetilde{\omega}_1,\widetilde{\omega}_2) + bg_{\mathbb{C}}(\widetilde{\omega}_2,\widetilde{\omega}_1) + abg_{\mathbb{C}}(\widetilde{\omega}_1,\widetilde{\omega}_1) \\
&= g_{\mathbb{C}}(\widetilde{\omega}_2,\widetilde{\omega}_3) + 2a
\end{aligned}$$

and

$$\begin{aligned}
0 &\overset{!}{=} g_{\mathbb{C}}(\omega_3,\omega_3) \\
&= g_{\mathbb{C}}(\widetilde{\omega}_3,\widetilde{\omega}_3) + 2ag_{\mathbb{C}}(\widetilde{\omega}_3,\widetilde{\omega}_2) + 2bg_{\mathbb{C}}(\widetilde{\omega}_3,\widetilde{\omega}_1) \\
&+ a^2 g_{\mathbb{C}}(\widetilde{\omega}_2,\widetilde{\omega}_2) + 2abg_{\mathbb{C}}(\widetilde{\omega}_2,\widetilde{\omega}_1) + b^2 g_{\mathbb{C}}(\widetilde{\omega}_1,\widetilde{\omega}_1) = g_{\mathbb{C}}(\widetilde{\omega}_3,\widetilde{\omega}_3) + 2b - 3a^2.
\end{aligned}$$

We note that with respect to $\omega_1,\omega_2,\omega_3$ the matrix of $\mathcal{W}$ is the same as before. We also note that

$$g_{\mathbb{C}}(\omega_1,\omega_3) = g_{\mathbb{C}}(\omega_2,\omega_2) = 1, \quad g_{\mathbb{C}}(\omega_1,\omega_1) = g_{\mathbb{C}}(\omega_1,\omega_2) = 0.$$

Therefore the matrix of $g_{\mathbb{C}}$ with respect to the new basis has the desired form. □

6.5. Principal null vectors and bivectors

Definition 6.34. Let V be a 4-dimensional real vector space with a non-degenerate inner product $\langle\cdot,\cdot\rangle$ with $\sigma = 1$. An element $\beta \in \Lambda^2(V) \setminus \{0\}$ is called *$g_{\mathbb{C}}$-null* if $g_{\mathbb{C}}(\beta,\beta) = 0$.

Remark 6.35. Let $\beta \in \Lambda^2(V) \setminus \{0\}$. We have

$$\begin{aligned}
\beta \text{ is } g_{\mathbb{C}}\text{-null} &\iff \mathrm{Re}\,(g_{\mathbb{C}}(\beta,\beta)) = \mathrm{Im}\,(g_{\mathbb{C}}(\beta,\beta)) = 0 \\
&\iff \langle\beta,\beta\rangle = \langle\beta,\star\beta\rangle = 0 \\
&\iff \beta \text{ is lightlike and decomposable}.
\end{aligned}$$

Let β be $g_{\mathbb{C}}$-null. By Lemma 6.25 we have $\beta = \ell \wedge x$, where ℓ is lightlike, x is spacelike and $\ell \perp x$. From Lemma 6.26 it follows that $\star\beta = \ell \wedge y$, where y is spacelike, $y \perp \ell$, $y \perp x$ and $|y| = |x|$. We have also seen that the line $\mathbb{R}\cdot\ell$ is uniquely determined by β, thus we write $N(\beta) := \mathbb{R}\cdot\ell$.

Definition 6.36. An element $\beta \in \Lambda^2(V) \setminus \{0\}$ is called a *principal null bivector* of $\mathcal{W}$ if $g_{\mathbb{C}}(\beta,\beta) = 0$ and $g_{\mathbb{C}}(\mathcal{W}\beta,\beta) = 0$.

Remark 6.37. If β is $g_{\mathbb{C}}$-null and an eigenvector of $\mathcal{W}$, then β is a principal null bivector, since $g_{\mathbb{C}}(\mathcal{W}\beta,\beta) = \lambda g_{\mathbb{C}}(\beta,\beta) = 0$. On the other hand we will soon find principal null bivectors which are not eigenvectors of $\mathcal{W}$.

Lemma 6.38. *Let β,β' be $g_{\mathbb{C}}$-null. Then we have $N(\beta) = N(\beta')$ if and only if there exists $z \in \mathbb{C}^*$ such that $\beta' = z\beta$.*

Proof. "$\Longleftarrow$" Let $\beta' = z\beta$, where $\beta = \ell \wedge x$, $z = a + ib$. Then

$$\beta' = (a+ib)\ell \wedge x = a\ell \wedge x + b \star (\ell \wedge x) = a\ell \wedge x + b\ell \wedge y = \ell \wedge (ax + by)$$

where $ax + by \perp \ell$ and $ax + by$ is spacelike. It follows that $N(\beta') = \mathbb{R} \cdot \ell = N(\beta)$.
"$\Longrightarrow$" Let $N(\beta) = N(\beta') = \mathbb{R} \cdot \ell$. Then we have $\beta = \ell \wedge x$, $\beta' = \ell \wedge x'$ and $\star\beta = \ell \wedge y$ where x, x', y are spacelike and $x, x', y \perp \ell$. The vectors x, y, ℓ form a basis of $\ell^{\perp}$ and thus there exist $a, b, c \in \mathbb{R}$ such that $x' = ax + by + c\ell$. It follows that $\beta' = a\ell \wedge x + b\ell \wedge y = (a + ib)\beta$. □

Hence we have a $1:1$-correspondence

$$\{\mathbb{C} \cdot \beta \,|\, \beta \in \Lambda^2(V) \text{ is } g_{\mathbb{C}}\text{-null}\} \longleftrightarrow \{\mathbb{R} \cdot \ell \,|\, \ell \in V \text{ is lightlike}\}$$

where the maps $\mathbb{C} \cdot \beta \mapsto N(\beta)$ and $\mathbb{R} \cdot \ell \mapsto \mathbb{C} \cdot (\ell \wedge x)$ for some spacelike $x \neq 0$, $x \perp \ell$, are inverse to each other.

Definition 6.39. **(1)** A vector $\ell \in V$ is called a *principal null vector* of W if there exists $x \in V$ such that $\ell \wedge x$ is a principal null bivector of $\mathcal{W}$. In this case the line $\mathbb{R} \cdot \ell$ is called a *principal null direction* of W.

(2) Let $\mathcal{W} \neq 0$. The *multiplicity of the principal null bivector* β is given by

$$m = \begin{cases} 1, & \text{if } \beta \text{ is not an eigenvector of } \mathcal{W}, \\ 2, & \text{if } \beta \text{ is an eigenvector of } \mathcal{W} \text{ for an eigenvalue } \lambda \neq 0, \\ 3, & \text{if } \beta \in \ker(\mathcal{W}) \text{ and } \dim_{\mathbb{C}} \ker(\mathcal{W}) = 1, \\ 4, & \text{if } \beta \in \ker(\mathcal{W}) \text{ and } \dim_{\mathbb{C}} \ker(\mathcal{W}) = 2. \end{cases}$$

The *multiplicity of a principal null vector* of W is defined as the multiplicity of the corresponding principal null bivector of $\mathcal{W}$.

Proposition 6.40. *Let $\mathcal{W} \neq 0$. Depending on the Petrov type, $\mathcal{W}$ has the following principal null bivectors (pnb's) and no further ones (up to multiplication by a complex number):*

type I:	*4 pnb's each with multiplicity $m = 1$;*
type II:	*1 pnb with $m = 2$ and 2 pnb's each with $m = 1$;*
type III:	*1 pnb with $m = 3$ and 1 pnb with $m = 1$;*
type D:	*2 pnb's each with $m = 2$;*
type N:	*1 pnb with $m = 4$.*

Note that the sum of the multiplicities equals 4 in each case. The list shows that the Petrov types can be characterized by in terms of the principal null bivectors of $\mathcal{W}$.

Proof. Let $\omega_1, \omega_2, \omega_3$ be a basis of $\Lambda^2(V)$ as in Lemma 6.33.
a) Consider type N. From Lemma 6.33 we see that $\beta := \omega_2$ is $g_{\mathbb{C}}$-null and $\beta \in \ker(\mathcal{W})$. For type N we have $\dim\ker(\mathcal{W}) = 2$ and thus $\beta = \omega_2$ is a principal null bivector with $m = 4$. Conversely, assume that $\beta = a\omega_1 + b\omega_2 + c\omega_3$ is a principal null bivector. Then we have

$$0 = g_{\mathbb{C}}(\beta,\beta) = a^2 + 2bc, \quad 0 = g_{\mathbb{C}}(\mathcal{W}\beta,\beta) = g_{\mathbb{C}}(c\omega_2,\beta) = c^2$$

and therefore $c = 0$, $a = 0$ and $\beta = b\omega_2$.
b) Consider type II. Let $\beta = a\omega_1 + b\omega_2 + c\omega_3$ and assume that β is a principal null bivector. Then we have

$$0 = g_{\mathbb{C}}(\beta,\beta) = a^2 + 2bc \text{ and}$$
$$0 = g_{\mathbb{C}}(\mathcal{W}\beta,\beta) = g_{\mathbb{C}}(a\lambda_1\omega_1 + b\lambda_2\omega_2 + c\lambda_3\omega_3 + c\omega_2, a\omega_1 + b\omega_2 + c\omega_3) = a^2\lambda_1 + 2\lambda_2 bc + c^2.$$

Substituting the first equation into the second one and using that $\lambda_1 = -2\lambda_2$ we get

$$0 = c^2 + 6\lambda_2 bc = c(c + 6\lambda_2 b).$$

Case 1: $c = 0$. Then we get $a = 0$ and $\beta = b\omega_2$ is a principal null bivector with $m = 2$.
Case 2: $c \neq 0$. Then $c = -6\lambda_2 b$ and thus $a^2 = -2bc = 12\lambda_2 b^2$. It follows that $a = \pm b\sqrt{12\lambda_2}$ and thus

$$\beta = \pm b\sqrt{12\lambda_2}\omega_1 + b\omega_2 - 6\lambda_2 b\omega_3.$$

This gives 2 linearly independent vectors since $\lambda_2 \neq 0$ and both have $m = 1$.
Conversely, every vector β of this form is a principal null bivector.
c) Consider type I. Assume that β is a principal null bivector. Then we get

$$0 = g_{\mathbb{C}}(\beta,\beta) = a^2 + b^2 + c^2,$$
$$0 = g_{\mathbb{C}}(\mathcal{W}\beta,\beta) = \lambda_1 a^2 + \lambda_2 b^2 + \lambda_3 c^2.$$

We multiply the first equation by λ_1 and subtract it from the second equation to get

$$0 = (\lambda_2 - \lambda_1)b^2 + (\lambda_3 - \lambda_1)c^2 \text{ and thus } b = \pm c\sqrt{\frac{\lambda_3 - \lambda_1}{\lambda_1 - \lambda_2}}.$$

We substitute this in the first equation and we get

$$0 = a^2 + \frac{\lambda_3 - \lambda_1}{\lambda_1 - \lambda_2}c^2 + c^2 = a^2 + c^2\frac{\lambda_3 - \lambda_2}{\lambda_1 - \lambda_2} \text{ and thus } a = \pm c\sqrt{\frac{\lambda_2 - \lambda_3}{\lambda_1 - \lambda_2}}.$$

Therefore we have

$$\beta = \pm c\sqrt{\frac{\lambda_2 - \lambda_3}{\lambda_1 - \lambda_2}}\omega_1 \pm c\sqrt{\frac{\lambda_3 - \lambda_1}{\lambda_1 - \lambda_2}}\omega_2 + c\omega_3$$

and this gives 4 vectors none of which is a complex multiple of another one since the λ_j are pairwise distinct. Conversely, every vector β of this form is a principal null bivector.
d) The assertion on types III, D is left as an exercise. □

Definition 6.41. Let $(V, \langle\cdot,\cdot\rangle)$ be a 4-dimensional Lorentzian vector space. An ordered basis (k, ℓ, x, y) of V is called a *real null tetrad* if the matrix of $\langle\cdot,\cdot\rangle$ with respect to (k, ℓ, x, y) is given by

$$\begin{pmatrix} 0 & -1 & & \\ -1 & 0 & & \\ & & 1 & 0 \\ & & 0 & 1 \end{pmatrix}.$$

In other words: k, ℓ are lightlike, $\langle k, \ell\rangle = -1$, x, y are spacelike and orthonormal and $\mathbb{R}\langle k, \ell\rangle \perp \mathbb{R}\langle x, y\rangle$.

Example 6.42. If e_0, e_1, e_2, e_3 is an orthonormal basis of V with $\varepsilon_0 = -1$, then

$$k := \frac{1}{\sqrt{2}}(e_0 + e_1), \quad \ell := \frac{1}{\sqrt{2}}(e_0 - e_1), \quad x := e_2, \quad y := e_3$$

is a real null tetrad.

Remark 6.43. **(1)** For every real null tetrad there exists an orthonormal basis $e_0, ..., e_3$ as in the example.
(2) Given a system of $1, 2$ or 3 vectors satisfying the relations of a real null tetrad, then this system can be completed to a real null tetrad.

(3) Let (k, ℓ, x, y) be a real null tetrad on V. Then $k \wedge \ell$, $k \wedge x$, $k \wedge y$, $\ell \wedge x$, $\ell \wedge y$, $x \wedge y$ form a real basis of $\Lambda^2(V)$ and the matrix of the induced inner product $\langle \cdot, \cdot \rangle$ on $\Lambda^2(V)$ with respect to this basis is given by

$$\begin{pmatrix} -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Proposition 6.44. *Let V be a 4-dimensional Lorentzian vector space and let $W \neq 0$ be an algebraic curvature tensor on V with $C(W) = 0$. Let $\ell \in V$ be lightlike.*

(a) ℓ is a principal null vector of W if and only if $W(\ell, x, \ell, y) = 0$ for all $x, y \perp \ell$.
(b) ℓ is a principal null vector of W of multiplicity $m \geq 2$ if and only if $W(\ell, x, \ell, y) = 0$ for all $x \perp \ell$ and for all $y \in V$.
(c) ℓ is a principal null vector of W of multiplicity $m \geq 3$ if and only if $W(\ell, x, \cdot, \cdot) = 0$ for all $x \perp \ell$.
(d) ℓ is a principal null vector of W of multiplicity $m \geq 4$ if and only if $W(\ell, \cdot, \cdot, \cdot) = 0$.

Proof. (a) "$\Longleftarrow$" Let x be spacelike, $x \neq 0$, $x \perp \ell$. Put $\beta := \ell \wedge x$. Then β is $g_{\mathbb{C}}$-null and $N(\beta) = \mathbb{R} \cdot \ell$. We write $\star\beta = \ell \wedge y$. Then by hypothesis we have

$$0 = W(\ell, x, \ell, x) = W(\ell \wedge x, \ell \wedge x) = \langle \mathcal{W}\beta, \beta \rangle = \mathrm{Re}(g_{\mathbb{C}}(\mathcal{W}\beta, \beta)),$$
$$0 = W(\ell, x, \ell, y) = W(\ell \wedge x, \ell \wedge y) = \langle \mathcal{W}\beta, \star\beta \rangle = -\mathrm{Im}(g_{\mathbb{C}}(\mathcal{W}\beta, \beta)).$$

Therefore we have $g_{\mathbb{C}}(\mathcal{W}\beta, \beta) = 0$ and thus β is a principal null bivector.
"$\Longrightarrow$" Let ℓ be a principal null vector of W. By definition there exists $x_0 \in V$ such that $\beta := \ell \wedge x_0$ is a principal null bivector of $\mathcal{W}$ and x_0 is spacelike, $x_0 \neq 0$, $x_0 \perp \ell$. We write $\star\beta = \ell \wedge y_0$. Then we have

$$0 = g_{\mathbb{C}}(\mathcal{W}\beta, \beta) = \langle \mathcal{W}\beta, \beta \rangle - i \langle \mathcal{W}\beta, \star\beta \rangle = W(\ell, x_0, \ell, x_0) - iW(\ell, x_0, \ell, y_0)$$

and thus $W(\ell, x_0, \ell, x_0) = 0 = W(\ell, x_0, \ell, y_0)$. Note that the vectors ℓ, x_0, y_0 form a basis of $\ell^\perp$. Therefore $W(\ell, x_0, \ell, y) = 0$ for all $y \perp \ell$. Now the vector $\star\beta = \ell \wedge y_0$ is also a principal null bivector of $\mathcal{W}$. Thus by the same argument we get $W(\ell, y_0, \ell, y) = 0$ for all $y \perp \ell$. Using again that the vectors ℓ, x_0, y_0 form a basis of $\ell^\perp$ we get $W(\ell, x, \ell, y) = 0$ for all $x, y \perp \ell$.
(b) "$\Longrightarrow$" Let ℓ be a principal null vector with $m \geq 2$. By definition there exists a spacelike $x_0 \in V$, $x_0 \perp \ell$, $x_0 \neq 0$ such that $\beta := \ell \wedge x_0$ is a principal null bivector and an eigenvector of $\mathcal{W}$. Then there exists a spacelike $y_0 \in V$, $y_0 \perp \ell$, $y_0 \perp x$, $|y_0| = |x_0|$ such that $\star\beta = \ell \wedge y_0$. Without

loss of generality we may assume $|x_0| = |y_0| = 1$. We extend the system ℓ, x_0, y_0 to a real null tetrad k, ℓ, x_0, y_0. Now let $y \in V$ and write $y = ak + b\ell + cx_0 + dy_0$. Using part (a) we get

$$\begin{aligned} W(\ell, x_0, \ell, y) &= aW(\ell, x_0, \ell, k) + cW(\ell, x_0, \ell, x_0) + dW(\ell, x_0, \ell, y_0) \\ &= aW(\ell, x_0, \ell, k) = a\langle \mathcal{W}\beta, \ell \wedge k\rangle = a\lambda\langle \ell \wedge x_0, \ell \wedge k\rangle = 0. \end{aligned}$$

In the same way we get $W(\ell, y_0, \ell, y) = 0$ and since the vectors ℓ, x_0, y_0 form a basis of $\ell^\perp$ we get $W(\ell, x, \ell, y) = 0$ for all $x \perp \ell$ and all $y \in V$.
"$\Longleftarrow$" Let x be spacelike, $x \neq 0$, $x \perp \ell$. Put $\beta := \ell \wedge x$ and write $\star\beta = \ell \wedge y$. By the proof of part (a), β is a principal null bivector. We extend the system ℓ, x, y to a real null tetrad k, ℓ, x, y. By hypothesis for all $z \in V$ we have

$$0 = W(\ell, x, \ell, z) = \langle \mathcal{W}\beta, \ell \wedge z\rangle \text{ and } 0 = W(\ell, y, \ell, z) = \langle \mathcal{W} \star \beta, \ell \wedge z\rangle$$

and therefore

$$\mathcal{W}\beta, \mathcal{W} \star \beta \in \mathbb{R}\langle \ell \wedge x, \ell \wedge y, k \wedge \ell\rangle^\perp = \mathbb{R}\langle \ell \wedge x, \ell \wedge y, x \wedge y\rangle.$$

It follows that

$$\langle \mathcal{W}\beta, x \wedge y\rangle = \pm\langle \mathcal{W}\beta, \star(k \wedge \ell)\rangle = \pm\langle \star\mathcal{W}\beta, k \wedge \ell\rangle = \pm\langle \mathcal{W} \star \beta, k \wedge \ell\rangle = 0$$

and therefore $\mathcal{W}\beta \in \mathbb{R}\langle \ell \wedge x, \ell \wedge y\rangle = \mathbb{R}\langle \beta, i\beta\rangle$ is an eigenvector of $\mathcal{W}$.
(c) By definition ℓ is a principal null vector of W if and only if there is a spacelike $x \in V \setminus \{0\}$, $x \perp \ell$ such that $\beta := \ell \wedge x$ is a principal null bivector of $\mathcal{W}$. Now we have

$$\begin{aligned} \mathcal{W}\beta = 0 &\Longleftrightarrow \langle \mathcal{W}\beta, \alpha\rangle = 0 \text{ for all } \alpha \in \Lambda^2(V) \\ &\Longleftrightarrow 0 = \langle \mathcal{W}\beta, u \wedge v\rangle = W(\ell, x, u, v) \text{ for all } u, v \in V. \end{aligned}$$

(d) "$\Longleftarrow$" Let $W(\ell, \cdot, \cdot, \cdot) = 0$. Then there is a spacelike $x \in V$, $|x| = 1$, $x \perp \ell$ such that $\beta = \ell \wedge x$ is a principal null bivector, $\beta \in \ker \mathcal{W}$. In particular we have $m \geq 3$. We extend the system ℓ, x to a real null tetrad ℓ, k, x, y. Then $\beta' := \ell \wedge k$ is not a complex multiple of β and we have

$$\langle \mathcal{W}\beta', u \wedge v\rangle = W(\ell, k, u, v) = 0$$

for all $u, v \in V$. It follows that $\mathcal{W}\beta' = 0$ and thus $\dim_{\mathbb{C}} \ker \mathcal{W} \geq 2$, i.e. $m \geq 4$.
"$\Longrightarrow$" Let $\beta = \ell \wedge x$ be a principal null bivector of $\mathcal{W}$ of multiplicity $m \geq 4$. By Proposition 6.40 we get that W has type N. Let $\omega_1, \omega_2, \omega_3$ be the basis of V from Lemma 6.33. Since $\beta \in \ker \mathcal{W}$ there exist $a, b \in \mathbb{R}$ such that $\beta = a\omega_1 + b\omega_2$. Since $g_{\mathbb{C}}(\beta, \beta) = 0$ we get $0 = a^2$ and thus $\beta = b\omega_2$. Let $\beta' \in \Lambda^2(V) \setminus \{0\}$ such that $\mathcal{W}\beta' = 0$, $g_{\mathbb{C}}(\beta, \beta') = 0$ and such that β, β' are linearly independent over $\mathbb{C}$. We extend the system ℓ, x to a real null tetrad k, ℓ, x, y. Then $k \wedge \ell, \ell \wedge x, k \wedge x$ form a complex basis of $\Lambda^2(V)$. We write

$$\beta' = z_1 k \wedge \ell + z_2 \ell \wedge x + z_3 k \wedge x, \quad z_1, z_2, z_3 \in \mathbb{C}.$$

We may assume that $z_2 = 0$, since otherwise we replace β' by $\beta' - z_2 \ell \wedge x$ and we still have that $\mathcal{W}\beta' = 0$, $g_{\mathbb{C}}(\beta, \beta') = 0$ and β, β' are linearly independent over $\mathbb{C}$. Using Remark 6.43 and using that $\star(\ell \wedge x) = \pm \ell \wedge y$ we get

$$\begin{aligned} 0 &= g_{\mathbb{C}}(\beta, \beta') = z_1 g_{\mathbb{C}}(k \wedge \ell, \ell \wedge x) + z_3 g_{\mathbb{C}}(k \wedge x, \ell \wedge x) \\ &= z_1 [\underbrace{\langle k \wedge \ell, \ell \wedge x\rangle}_{=0} \pm i \underbrace{\langle k \wedge \ell, \ell \wedge y\rangle}_{=0}] + z_3 [\underbrace{\langle k \wedge x, \ell \wedge x\rangle}_{=-1} \pm i \underbrace{\langle k \wedge x, \ell \wedge y\rangle}_{=0}] = -z_3. \end{aligned}$$

Thus we may assume that $\beta' = k \wedge \ell$. By part (c) we have $W(\ell, u, \cdot, \cdot) = 0$ for all $u \perp \ell$, i.e. for all $u \in \mathbb{R}\langle \ell, x, y\rangle$. In addition we have

$$W(\ell, k, \cdot, \cdot) = \langle \mathcal{W}(\ell \wedge k), \cdot \wedge \cdot \rangle = -\langle \mathcal{W}\beta', \cdot \wedge \cdot \rangle = 0$$

and thus $W(\ell, \cdot, \cdot, \cdot) = 0$. $\square$

Corollary 6.45 (Bel-Sachs criteria). *Let $W \neq 0$ be an algebraic curvature tensor on a 4-dimensional Lorentzian vector space V. Then we have*

$$\begin{aligned} W \text{ has type } N &\iff \exists \ell \in V \text{ lightlike such that } W(\ell, \cdot, \cdot, \cdot) = 0; \\ W \text{ has type } III &\iff W \text{ does not have type } N \text{ and } \exists \ell \in V \text{ lightlike such that } W(\ell, x, \cdot, \cdot) = 0 \\ &\qquad \text{for all } x \perp \ell; \\ &\iff \exists \ell_1, \ell_2 \in V \text{ lightlike and linearly independent such that} \\ &\qquad W(\ell_1, x, \cdot, \cdot) = 0 \text{ for all } x \perp \ell_1 \text{ and } W(\ell_2, x, \ell_2, y) = 0 \text{ for all } x, y \perp \ell_2; \\ W \text{ has type } D &\iff \exists \ell_1, \ell_2 \in V \text{ lightlike and linearly independent such that} \\ &\qquad W(\ell_j, x, \ell_j, \cdot) = 0 \text{ for all } x \perp \ell_j,\ j = 1, 2; \\ W \text{ has type } II &\iff \exists \ell_1, \ell_2, \ell_3 \in V \text{ lightlike and pairwise linearly independent such that} \\ &\qquad W(\ell_1, x, \ell_1, \cdot) = 0 \text{ for all } x \perp \ell_1 \text{ and } W(\ell_j, x, \ell_j, y) = 0 \text{ for all } x, y \perp \ell_j, \\ &\qquad j = 2, 3; \\ W \text{ has type } I &\iff \exists \ell_1, \ell_2, \ell_3, \ell_4 \in V \text{ lightlike and pairwise linearly independent such that} \\ &\qquad W(\ell_j, x, \ell_j, y) = 0 \text{ for all } x, y \perp \ell_j,\ j = 1, ..., 4. \end{aligned}$$

6.6. The optical scalars

We want to apply the results of the previous section in order to study lightlike geodesics on a spacetime M.

Definition 6.46. A vector field X on a semi-Riemannian manifold M is called a *geodesic vector field* if its integral curves are geodesics, i.e. if $\nabla_X X = 0$.

Let ℓ be a geodesic lightlike vector field. Since $\langle \ell, \ell \rangle \equiv 0$ we get for all $X \in TM$

$$0 = \partial_X \langle \ell, \ell \rangle = 2 \langle \nabla_X \ell, \ell \rangle \text{ and thus } \nabla_X \ell \perp \ell.$$

Therefore the map $X \mapsto \nabla_X \ell$ restricts to the map

$$\ell^\perp \to \ell^\perp, \quad X \mapsto \nabla_X \ell.$$

At every point of M we have $\ell \in \ell^\perp$ and we consider the quotient vector space $\ell^\perp/\ell$. Since we have $\nabla_\ell \ell = 0$ the above map induces a well defined map

$$\begin{tikzcd} \ell^\perp/\ell \arrow[r] \arrow[dr, "D"'] & \ell^\perp \arrow[d] \\ & \ell^\perp/\ell \end{tikzcd} \qquad \begin{tikzcd} {[X]} \arrow[r, mapsto] & \nabla_X \ell \arrow[d, mapsto] \\ & {[\nabla_X \ell]} \end{tikzcd} .$$

Here the brackets $[\cdot]$ denote the equivalence class modulo ℓ.
At every point of M the vector space $\ell^\perp/\ell$ has real dimension 2. The Lorentzian metric restricted to $\ell^\perp$ is positive semidefinite and degenerate and ℓ is precisely the null space. Thus we get an induced metric $\langle \cdot, \cdot \rangle$ on $\ell^\perp/\ell$ which is positive definite. Every algebraic complement of $\mathbb{R}\langle \ell \rangle$ in $\ell^\perp$ is isomorphic to $\ell^\perp/\ell$. In particular, there is no canonical way to identify $\ell^\perp/\ell$ with a linear subspace of $\ell^\perp$.

Definition 6.47. The vector bundle on M with fibers $\ell^\perp/\ell$ is called the *screen bundle* on M.

Let $[x], [y]$ be an orthonormal basis of $\ell^\perp/\ell$. With respect to this basis the map D defined above has the matrix

$$\begin{pmatrix} \ell_{x,x} & \ell_{y,x} \\ \ell_{x,y} & \ell_{y,y} \end{pmatrix}$$

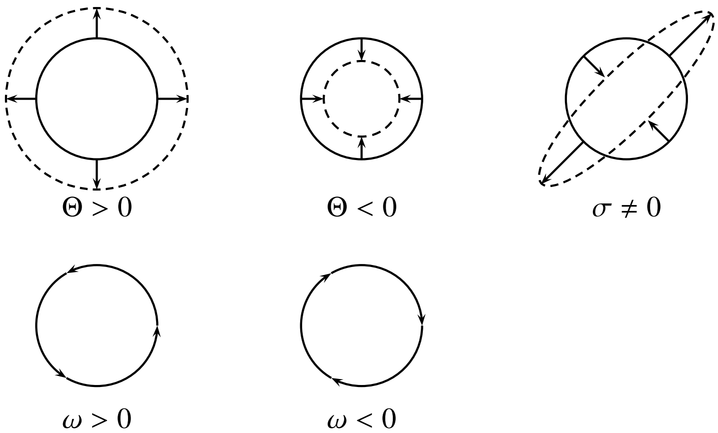
where $\ell_{y,x} := \langle \nabla_x \ell, y \rangle$. We define D^t as the adjoint endomorphism of D with respect to $\langle \cdot, \cdot \rangle$. Now we decompose D into a symmetric and an antisymmetric part and we decompose the symmetric part further into a term with vanishing trace and a multiple of the identity, i.e.

$$\begin{aligned} D &= \tfrac{1}{2}(D + D^t) + \tfrac{1}{2}(D - D^t) \\ &= \tfrac{1}{2}\mathrm{tr}(D)\mathrm{id} + \tfrac{1}{2}(D + D^t - \mathrm{tr}(D)\mathrm{id}) + \tfrac{1}{2}(D - D^t) \\ &= \begin{pmatrix} \theta & 0 \\ 0 & \theta \end{pmatrix} + \frac{1}{2} \begin{pmatrix} \ell_{x,x} - \ell_{y,y} & \ell_{x,y} + \ell_{y,x} \\ \ell_{x,y} + \ell_{y,x} & \ell_{y,y} - \ell_{x,x} \end{pmatrix} + \begin{pmatrix} 0 & \omega \\ -\omega & 0 \end{pmatrix} \\ &= \begin{pmatrix} \theta & 0 \\ 0 & \theta \end{pmatrix} + \frac{1}{2} \begin{pmatrix} -\mathrm{Re}\sigma & \mathrm{Im}\sigma \\ \mathrm{Im}\sigma & \mathrm{Re}\sigma \end{pmatrix} + \begin{pmatrix} 0 & \omega \\ -\omega & 0 \end{pmatrix} \end{aligned}$$

where we have defined

$$\begin{aligned}
\theta &:= \tfrac{1}{2}\mathrm{tr}(D) = \tfrac{1}{2}(\ell_{x,x} + \ell_{y,y}) \quad (\textit{expansion})\\
\omega &:= \tfrac{1}{2}(\ell_{y,x} - \ell_{x,y}) \quad (\textit{rotation} \text{ or } \textit{twist})\\
\sigma &:= \tfrac{1}{2}(\ell_{y,y} - \ell_{x,x}) + \frac{i}{2}(\ell_{y,x} + \ell_{x,y}) \quad (\textit{complex shear}).
\end{aligned}$$

The functions θ, ω, σ are called the *optical scalars* of ℓ. We note that θ is independent of the choice of basis of $\ell^\perp/\ell$ since θ is the trace of an endomorphism of $\ell^\perp/\ell$. Moreover since $\det(\frac{1}{2}(D - D^t)) = \omega^2$ we conclude that ω^2 is independent of the choice of basis. If we equip $\ell^\perp/\ell$ with an orientation and if we consider only positively oriented orthonormal bases $[x], [y]$ then the sign of ω is fixed and hence ω is also independent of the choice of basis.
We can interpret the optical scalars geometrically as follows. At some point p of M we identify $\ell^\perp/\ell$ with a vector subspace of $\ell^\perp$. Then for all points q on the integral curve of ℓ through p we regard the endomorphism D of $\ell^\perp/\ell$ at q as an endomorphism of $\ell^\perp/\ell$ at p via parallel transport along the integral curve. The optical scalars describe infinitesimal deformations of a small disk in $\ell^\perp/\ell$ at p. The number θ describes expansion or shrinking of the disc, ω describes rotation of the disc and σ describes the deformation of the disc into an ellipse of the same area.



Definition 6.48. Let M be a smooth manifold. A vector subbundle V of TM is called *integrable* if for every $x \in M$ there exists a submanifold Q of M such that $x \in Q$ and such that for all $p \in Q$ we have $T_pQ = V_p$.

By Frobenius' theorem V is integrable if and only if for all vector fields X, Y on M which are sections of V the vector field $[X, Y]$ is also a section of V.

Lemma 6.49. *The vector subbundle $\ell^\perp$ of TM is integrable if and only if $\omega = 0$.*

Proof. The vector fields ℓ, x, y form a basis of $\ell^\perp$ at every point of M, where $[x], [y]$ form an orthonormal basis of $\ell^\perp/\ell$. We get

$$\langle \nabla_x \ell, y\rangle = \partial_x \langle \ell, y\rangle - \langle \ell, \nabla_x y\rangle = -\langle \ell, \nabla_x y\rangle$$

and similarly $\left\langle \nabla_y \ell, x\right\rangle = -\left\langle \ell, \nabla_y x\right\rangle$. Thus we obtain

$$\begin{aligned}\omega = 0 &\Longleftrightarrow \langle \nabla_x \ell, y\rangle = \left\langle \nabla_y \ell, x\right\rangle \Longleftrightarrow \langle \ell, \nabla_x y\rangle = \left\langle \ell, \nabla_y x\right\rangle \\ &\Longleftrightarrow 0 = \left\langle \ell, \nabla_x y - \nabla_y x\right\rangle = \langle \ell, [x, y]\rangle \Longleftrightarrow [x, y] \in \ell^\perp.\end{aligned}$$

We always have $[\ell, x], [\ell, y] \in \ell^\perp$. Namely since ℓ is lightlike and geodesic, $\ell \perp x$, we get

$$\langle \ell, [\ell, x]\rangle = \langle \ell, \nabla_\ell x - \nabla_x \ell\rangle = \partial_\ell \langle \ell, x\rangle - \langle \nabla_\ell \ell, x\rangle - \tfrac{1}{2}\partial_x \langle \ell, \ell\rangle = 0$$

and similarly for $\langle \ell, [\ell, y]\rangle$. The assertion then follows from Frobenius' theorem. □

For an endomorphism A of $\ell^\perp/\ell$ we write $|A|^2 := \mathrm{tr}(A^t A)$. In the following we also write

$$D_s := \tfrac{1}{2}(D + D^t), \quad D_{s,0} := \tfrac{1}{2}(D + D^t - \mathrm{tr}(D)\mathrm{id}), \quad D_a := \tfrac{1}{2}(D - D^t).$$

We note that $|\sigma|^2$ is independent of the choice of basis. Namely we have

$$\begin{aligned}|\sigma|^2 &= (\mathrm{Re}\sigma)^2 + (\mathrm{Im}\sigma)^2 = \tfrac{1}{2}|D_{s,0}|^2 = \tfrac{1}{2}\left|\tfrac{1}{2}(D + D^t - \mathrm{tr}(D)\mathrm{id})\right|^2 = \frac{1}{8}\mathrm{tr}((D + D^t - \mathrm{tr}(D)\mathrm{id})^2) \\ &= \frac{1}{8}\mathrm{tr}(D^2 + (D^t)^2 + DD^t + D^t D - 2\mathrm{tr}(D)(D + D^t) + \mathrm{tr}(D)^2\mathrm{id}) \\ &= \frac{1}{8}(2\mathrm{tr}(D^2) + 2|D|^2 - 4\mathrm{tr}(D)^2 + 2\mathrm{tr}(D)^2) = \frac{1}{4}(\mathrm{tr}(D^2) + |D|^2 - \mathrm{tr}(D)^2) \qquad (6)\end{aligned}$$

and thus $|\sigma|^2$ is independent of the choice of basis. We note however that σ itself does depend on the choice of basis.

Next we note that the differential operator ∇_ℓ acting on vector fields on M induces differential operators on sections of both $\ell^\perp$ and $\ell^\perp/\ell$ which we also denote by ∇_ℓ. Namely if $X \perp \ell$, then we have

$$0 = \partial_\ell \langle X, \ell\rangle = \langle \nabla_\ell X, \ell\rangle + \langle X, \nabla_\ell \ell\rangle = \langle \nabla_\ell X, \ell\rangle$$

and therefore $\nabla_\ell X \in \ell^\perp$. It follows that ∇_ℓ defines a first order differential operator acting on sections of $\ell^\perp$. Furthermore for all vector fields X on M and for all $f \in C^\infty(M)$ we have

$$\nabla_\ell(X + f\ell) = \nabla_\ell X + (\partial_\ell f)\ell + f\nabla_\ell \ell = \nabla_\ell X + (\partial_\ell f)\ell.$$

It follows that ∇_ℓ induces a first order differential operator on sections of $\ell^\perp/\ell$.

For an endomorphism A of $\ell^\perp/\ell$ we define

$$\nabla_\ell A := \nabla_\ell \circ A - A \circ \nabla_\ell.$$

Then for all sections X of $\ell^\perp/\ell$ the Leibniz rule holds

$$\nabla_\ell(A(X)) = (\nabla_\ell A)(X) + A(\nabla_\ell X).$$

We also note that the endomorphism $X \mapsto R(\ell, X)\ell$ of TM induces an endomorphism $R(\ell, \cdot)\ell$ of $\ell^\perp/\ell$.

Proposition 6.50 (Riccati equation). *We have*

$$\nabla_\ell D + D^2 = R(\ell, \cdot)\ell.$$

Proof. Let $X \in T_pM$ for some $p \in M$ such that $X \perp \ell$ at p. We obtain a vector field X along the integral curve γ of ℓ through p by parallel translation along γ. In particular we may assume that $\nabla_\ell X = 0$. We get

$$\begin{aligned}(\nabla_\ell D)(X) &= \nabla_\ell(D(X)) - D(\underbrace{\nabla_\ell X}_{=0}) = \nabla_\ell\nabla_X\ell = R(\ell, X)\ell + \nabla_X \underbrace{\nabla_\ell\ell}_{=0} + \nabla_{[\ell,X]}\ell \\ &= R(\ell, X)\ell - \nabla_{D(X)}\ell = R(\ell, X)\ell - D^2(X),\end{aligned}$$

where we have used that $[\ell, X] = \nabla_\ell X - \nabla_X\ell = -\nabla_X\ell = -D(X)$. □

The following proposition tells us how the optical scalars evolve along the integral curves of ℓ.

Proposition 6.51 (Sachs equations). *The following equations hold:*

(i) $\partial_\ell\omega = -2\theta\omega$
(ii) $\partial_\ell\theta = \omega^2 - \theta^2 - |\sigma|^2 - \frac{1}{2}\mathrm{Ric}(\ell, \ell)$
(iii) $\partial_\ell\sigma = -2\theta\sigma - \frac{1}{2}(\langle R(\ell, X)\ell, X\rangle - \langle R(\ell, Y)\ell, Y\rangle) + i\,\langle R(\ell, X)\ell, Y\rangle$,
where σ is defined with respect to an orthonormal frame X, Y such that X, Y are parallel along the integral curves of ℓ.

Proof. (i) *Case 1: D_a is invertible everywhere.* By definition we have $\omega^2 = \det(D_a)$ and the formula for the derivative of the determinant yields

$$\partial_\ell\omega^2 = \omega^2 \mathrm{tr}(D_a^{-1}\nabla_\ell D_a). \tag{7}$$

In the following computation we use the Riccati equation and its transpose where we note that $\nabla_\ell D^t = (\nabla_\ell D)^t$ and that $X \mapsto R(\ell, X)\ell$ is self-adjoint with respect to $\langle\cdot,\cdot\rangle$. We obtain

$$\nabla_\ell D_a = \tfrac{1}{2}\nabla_\ell(D - D^t) = \tfrac{1}{2}(-D^2 + R(\ell,\cdot)\ell + (D^t)^2 - R(\ell,\cdot)\ell) = \tfrac{1}{2}((D^t)^2 - D^2).$$

We conclude that

$$\begin{aligned}
\operatorname{tr}(D_a^{-1}\nabla_\ell D_a) &= \tfrac{1}{2}\operatorname{tr}(D_a^{-1}((D^t)^2 - D^2)) = \tfrac{1}{2}\operatorname{tr}(D_a^{-1}((\underbrace{D^t - D}_{=-2D_a})(D^t + D) + DD^t - D^tD)) \\
&= -\operatorname{tr}(D^t + D) + \tfrac{1}{2}\operatorname{tr}(D_a^{-1}[D, \underbrace{D^t - D}_{=-2D_a}]) = -2\operatorname{tr}(D) - \operatorname{tr}(D_a^{-1}(DD_a - D_aD)) \\
&= -2\operatorname{tr}(D) - \operatorname{tr}(D) + \operatorname{tr}(D) = -2\operatorname{tr}(D) = -4\theta. \qquad (8)
\end{aligned}$$

Now equations (8) and (7) yield

$$2\omega\partial_\ell\omega = \partial_\ell\omega^2 = \omega^2\cdot(-4\theta)$$

which is assertion (i) in Case 1.
Case 2: D_a is not everywhere invertible. If I denotes the parameter interval, we put

$$I^+ := \{s \in I \,|\, D_a(s) \text{ is invertible}\}.$$

We have seen that (i) holds for all $s \in I^+$. By continuity (i) holds for all $s \in \overline{I^+}$. On $I \setminus \overline{I^+}$ we have $\omega^2 \equiv 0$ and thus $\partial_\ell\omega \equiv 0$. Therefore (i) also holds on $I \setminus \overline{I^+}$.
(ii) Using the Riccati equation we compute

$$\partial_\ell\theta = \tfrac{1}{2}\operatorname{tr}(\nabla_\ell D) = \tfrac{1}{2}\operatorname{tr}(-D^2 + R(\ell,\cdot)\ell) = -\tfrac{1}{2}\operatorname{tr}(D^2) + \tfrac{1}{2}\operatorname{tr}(R(\ell,\cdot)\ell).$$

If $(b_i)_{i=1}^4$ is a basis of T_pM and $(g_{ij}) := (\left\langle b_i, b_j\right\rangle)_{i,j=1}^4$ with inverse matrix (g^{ij}) then we have

$$\operatorname{Ric}(\ell,\ell) = \sum_{i,j=1}^4 g^{ij}\left\langle R(\ell, b_i)b_j, \ell\right\rangle.$$

We choose a real null tetrad ℓ, k, x, y as a basis of T_pM. Then $[x], [y]$ form an orthonormal basis of $\ell^\perp/\ell$ and we get

$$\operatorname{Ric}(\ell,\ell) = \langle R(\ell,x)x,\ell\rangle + \langle R(\ell,y)y,\ell\rangle - \underbrace{\langle R(\ell,\ell)k,\ell\rangle}_{=0} - \underbrace{\langle R(\ell,k)\ell,\ell\rangle}_{=0} = -\operatorname{tr}(R(\ell,\cdot)\ell)$$

and therefore

$$\partial_\ell\theta = -\tfrac{1}{2}\operatorname{tr}(D^2) - \tfrac{1}{2}\operatorname{Ric}(\ell,\ell). \qquad (9)$$

From the characteristic polynomial

$$\chi_{D_a}(\lambda) = \lambda^2 - \operatorname{tr}(D_a)\lambda + \det(D_a) = \lambda^2 + \det(D_a)$$

we get by Cayley-Hamilton's theorem that $0 = D_a^2 + \det(D_a)\mathrm{id}$ and thus

$$0 = \operatorname{tr}(D_a^2) + 2\det(D_a).$$

Using this and (6) we conclude that

$$
\begin{aligned}
\omega^2 - \theta^2 - |\sigma|^2 &= \det(D_a) - \frac{1}{4}\mathrm{tr}(D)^2 - \frac{1}{4}(\mathrm{tr}(D^2) - \mathrm{tr}(D)^2 + |D|^2) \\
&= -\tfrac{1}{2}\mathrm{tr}(D_a^2) - \frac{1}{4}\mathrm{tr}(D^2) - \frac{1}{4}|D|^2 \\
&= -\frac{1}{8}\mathrm{tr}(D^2 + (D^t)^2 - DD^t - D^tD) - \frac{1}{4}\mathrm{tr}(D^2) - \frac{1}{4}|D|^2 \\
&= -\frac{1}{4}\mathrm{tr}(D^2) + \frac{1}{4}|D|^2 - \frac{1}{4}\mathrm{tr}(D^2) - \frac{1}{4}|D|^2 \\
&= -\tfrac{1}{2}\mathrm{tr}(D^2)
\end{aligned}
$$

and plugging this term into (9) we obtain the assertion (*ii*).
(*iii*) Using the Riccati equation and its transpose we compute

$$\nabla_\ell D_{s,0} = \tfrac{1}{2}\nabla_\ell(D + D^t - \mathrm{tr}(D)\mathrm{id}) = \tfrac{1}{2}(-D^2 + 2R(\ell,\cdot)\ell - (D^t)^2 - (\partial_\ell\theta)\mathrm{id}).$$

Recall that with respect to the basis X, Y we have

$$D_{s,0} = \begin{pmatrix} -\mathrm{Re}\sigma & \mathrm{Im}\sigma \\ \mathrm{Im}\sigma & \mathrm{Re}\sigma \end{pmatrix}.$$

Since X, Y are orthonormal and parallel along the integral curves of ℓ we get

$$
\begin{aligned}
\partial_\ell\sigma &= \partial_\ell\Big(\tfrac{1}{2}\mathrm{Re}\sigma + \tfrac{1}{2}\mathrm{Re}\sigma + i\mathrm{Im}\sigma\Big) = \partial_\ell\Big(-\tfrac{1}{2}\langle D_{s,0}X, X\rangle + \tfrac{1}{2}\langle D_{s,0}Y, Y\rangle + i\langle D_{s,0}X, Y\rangle\Big) \\
&= -\tfrac{1}{2}\langle(\nabla_\ell D_{s,0})X, X\rangle + \tfrac{1}{2}\langle(\nabla_\ell D_{s,0})Y, Y\rangle + i\langle(\nabla_\ell D_{s,0})X, Y\rangle \\
&= -\frac{1}{4}(-\left\langle D^2X, X\right\rangle - \left\langle(D^t)^2X, X\right\rangle + 2\langle R(\ell, X)\ell, X\rangle - \partial_\ell\theta\langle X, X\rangle) \\
&+ \frac{1}{4}(-\left\langle D^2Y, Y\right\rangle - \left\langle(D^t)^2Y, Y\right\rangle + 2\langle R(\ell, Y)\ell, Y\rangle - \partial_\ell\theta\langle Y, Y\rangle) \\
&+ \frac{i}{2}(-\left\langle D^2X, Y\right\rangle - \left\langle(D^t)^2X, Y\right\rangle + 2\langle R(\ell, X)\ell, Y\rangle - \partial_\ell\theta\langle X, Y\rangle) \\
&= \underbrace{\tfrac{1}{2}(\left\langle D^2X, X\right\rangle - \left\langle D^2Y, Y\right\rangle) - \frac{i}{2}(\left\langle D^2X, Y\right\rangle + \left\langle X, D^2Y\right\rangle)}_{=:A} \\
&- \tfrac{1}{2}(\langle R(\ell, X)\ell, X\rangle - \langle R(\ell, Y)\ell, Y\rangle) + i\langle R(\ell, X)\ell, Y\rangle. \qquad (10)
\end{aligned}
$$

From the characteristic polynomial

$$\chi_D(\lambda) = \lambda^2 - \mathrm{tr}(D)\lambda + \det(D)$$

we get by Cayley-Hamilton's theorem that $0 = D^2 - \mathrm{tr}(D)D + \det(D)\mathrm{id}$ and thus

$$
\begin{aligned}
A &= \tfrac{1}{2}\langle(\mathrm{tr}(D)D - \det(D))X, X\rangle - \tfrac{1}{2}\langle(\mathrm{tr}(D)D - \det(D))Y, Y\rangle \\
&- \frac{i}{2}\langle(\mathrm{tr}(D)D - \det(D))X, Y\rangle - \frac{i}{2}\langle(\mathrm{tr}(D)D - \det(D))Y, X\rangle \\
&= \tfrac{1}{2}\mathrm{tr}(D)(\langle DX, X\rangle - \langle DY, Y\rangle - i\langle DX, Y\rangle - i\langle DY, X\rangle).
\end{aligned}
$$

Since $\langle DX, X\rangle = \langle D_sX, X\rangle$, $\langle DY, Y\rangle = \langle D_sY, Y\rangle$, $\langle X, X\rangle = \langle Y, Y\rangle$ and $\langle X, Y\rangle = 0$ we get

$$\begin{aligned} A &= \theta(\langle D_sX, X\rangle - \langle D_sY, Y\rangle - 2i\,\langle D_sX, Y\rangle) \\ &= \theta(\underbrace{\langle D_{s,0}X, X\rangle}_{=-\mathrm{Re}\sigma} - \underbrace{\langle D_{s,0}Y, Y\rangle}_{=\mathrm{Re}\sigma} - 2i\,\underbrace{\langle D_{s,0}X, Y\rangle}_{=\mathrm{Im}\sigma}) = -2\theta\sigma \end{aligned}$$

and by plugging this term into (10) we obtain the assertion (*iii*). □

Example 6.52. (a) Let $M = \mathbb{R}^4$ with coordinates (t, x, y, z) and Minkowski metric such that ∂_t is timelike. Then $\ell := \partial_t - \partial_x$ is a lightlike geodesic vector field. In this case ℓ is parallel and thus $D = 0$. It follows that $\omega = \theta = \sigma = 0$.

(b) Let $M := \mathbb{R}^4 \setminus (\mathbb{R} \times (0, 0, 0))$ with Minkowski metric and define $\ell := \partial_t + \partial_r$, where

$$r := (x^2 + y^2 + z^2)^{1/2}, \quad \partial_r := \frac{1}{r}(x\partial_x + y\partial_y + z\partial_z).$$

The integral curves of ℓ are straight lines on the lightcone. In particular, ℓ is a lightlike and geodesic vector field. In order to compute D we calculate the derivative $\nabla_Y \ell$ where $Y \perp \ell$. With respect to the coordinates on $\mathbb{R} \times \mathbb{R}^3$ we write $Y = (\tau, \vec{y})$, $\tau \in \mathbb{R}$, $\vec{y} \in \mathbb{R}^3$ and $\ell = (1, \frac{\vec{x}}{r})$. The condition $Y \perp \ell$ leads to

$$0 = \langle Y, \ell\rangle = -\tau + \langle \vec{y}, \partial_r\rangle, \qquad \text{and thus } \tau = \langle \vec{y}, \partial_r\rangle.$$

We conclude that

$$\nabla_Y \ell = \nabla_Y \partial_r = \nabla_{\vec{y}} \frac{\vec{x}}{r} = -\frac{1}{r^2}\langle \vec{y}, \nabla r\rangle\, \vec{x} + \frac{1}{r}\nabla_{\vec{y}}\vec{x} = -\frac{1}{r^2}\left\langle \vec{y}, \frac{\vec{x}}{r}\right\rangle \vec{x} + \frac{1}{r}\vec{y} = -\frac{1}{r^2}\langle \vec{y}, \vec{x}\rangle\, \frac{\vec{x}}{r} + \frac{1}{r}\vec{y}.$$

Since $\ell = (1, \frac{\vec{x}}{r})$ we get modulo ℓ

$$\nabla_Y \ell \overset{\mathrm{mod}\ \mathbb{R}\ell}{=} \frac{1}{r^2}\langle \vec{x}, \vec{y}\rangle\, \partial_t + \frac{1}{r}\vec{y} = \frac{1}{r}\Big(\left\langle \vec{y}, \frac{\vec{x}}{r}\right\rangle \partial_t + \vec{y}\Big) = \frac{1}{r}Y$$

It follows that $D = \frac{1}{r}\mathrm{id}$ and thus $\theta = \frac{1}{r}$, $\omega = 0$, $\sigma = 0$.

(c) Let $M = \mathrm{Kerr}_{m,a}$ equipped with Boyer-Lindquist coordinates and define

$$\ell = \partial_r + \frac{1}{\Delta}V = \frac{\rho}{\sqrt{\varepsilon\Delta}}E_1 + \frac{\varepsilon\rho}{\sqrt{\varepsilon\Delta}}E_4 = \frac{\rho}{\sqrt{\varepsilon\Delta}}(E_1 + \varepsilon E_4),$$

where $\varepsilon = \mathrm{sign}\Delta$, $V = (r^2 + a^2)\partial_t + a\partial_\varphi$ and we use the orthonormal frame E_1, E_2, E_3, E_4 defined in (**??**). Then ℓ is a lightlike geodesic vector field. The vectors $[E_2], [E_3]$ form an orthonormal basis of $\ell^\perp/\ell$. Using the terms $\nabla_{E_i}E_j$ from Section 4.2.8 we compute modulo

$\mathbb{R}\ell$

$$\begin{aligned}
\nabla_{E_2}\ell &= \partial_{E_2}\Big(\frac{\rho}{\sqrt{\varepsilon\Delta}}\Big)\underbrace{(E_1+\varepsilon E_4)}_{\in\mathbb{R}\ell} + \frac{\rho}{\sqrt{\varepsilon\Delta}}(\nabla_{E_2}E_1+\varepsilon\nabla_{E_2}E_4)\\
&\overset{\mathrm{mod}\,\mathbb{R}\ell}{=} \frac{\rho}{\sqrt{\varepsilon\Delta}}\Big(\frac{r\sqrt{\varepsilon\Delta}}{\rho^3}E_2+\varepsilon\frac{\varepsilon a\cos\vartheta\sqrt{\varepsilon\Delta}}{\rho^3}E_3\Big) = \frac{1}{\rho^2}(rE_2+a\cos\vartheta E_3)\\
\nabla_{E_3}\ell &\overset{\mathrm{mod}\,\mathbb{R}\ell}{=} \frac{\rho}{\sqrt{\varepsilon\Delta}}(\nabla_{E_3}E_1+\varepsilon\nabla_{E_3}E_4)\\
&= \frac{\rho}{\sqrt{\varepsilon\Delta}}\Big(\frac{r\sqrt{\varepsilon\Delta}}{\rho^3}E_3-\frac{\varepsilon ar\sin\vartheta}{\rho^3}E_4-\frac{ar\sin\vartheta}{\rho^3}E_1-\frac{a\cos\vartheta\sqrt{\varepsilon\Delta}}{\rho^3}E_2\Big)\\
&\overset{\mathrm{mod}\,\mathbb{R}\ell}{=} \frac{1}{\rho^2}(rE_3-a\cos\vartheta E_2).
\end{aligned}$$

Therefore the matrix of D with respect to the basis $[E_2], [E_3]$ is given by

$$\frac{1}{\rho^2}\begin{pmatrix} r & -a\cos\vartheta \\ a\cos\vartheta & r\end{pmatrix}$$

and thus $\theta = \frac{r}{\rho^2}$, $\omega = -\frac{a\cos\vartheta}{\rho^2}$, $\sigma = 0$. Note that the rotation ω changes its sign at the equatorial plane Eq. In the special case of the Schwarzschild metric we have $a = 0$ thus $\rho = r$ and therefore $\theta = \frac{1}{r}$, $\omega = 0$, $\sigma = 0$ as in Minkowski space.

Remark 6.53. If we define $\widehat{\rho} := -\theta + i\omega$ then the equations (i), (ii) of Proposition 6.51 are equivalent to

$$\partial_\ell\widehat{\rho} = -\partial_\ell\theta + i\partial_\ell\omega = -2i\theta\omega - \omega^2 + \theta^2 + |\sigma|^2 + \tfrac{1}{2}\mathrm{Ric}(\ell,\ell) = \widehat{\rho}^2 + |\sigma|^2 + \tfrac{1}{2}\mathrm{Ric}(\ell,\ell).$$

This leads to the so-called Newman-Penrose formalism where one uses complex null tetrads.

Theorem 6.54 (Goldberg-Sachs). *Let M be a 4-dimensional Lorentzian manifold with* $\mathrm{Ric} = 0$ *and $R \neq 0$ and let ℓ be a lightlike vector field on M. Then the following are equivalent.*

(i) There exists $f \in C^\infty(M)$, $f > 0$ such that $f\ell$ is a geodesic vector field with $\sigma = 0$.
(ii) At every point ℓ is a principal null vector of R with multiplicity $m \geq 2$.

Corollary 6.55. *If M has Petrov type I, then M has no lightlike geodesic vector fields with $\sigma = 0$.*

A. Solutions of selected exercises

1.1. We write $x(t)$ for the distance from space craft to earth at time t. From $x(0) = 0$, $\dot{x}(0) = 0$ and $\ddot{x} = g$, we get

$$x(t) = \tfrac{1}{2}gt^2$$

for the first half of the journey. Is D the distance between earth and the object X and T is the total time of travel, we obtain $D/2 = \frac{1}{2}g(T/2)^2$ and therefore

$$T = 2\sqrt{D/g}.$$

The maximal velocity $v_{\max}$ is achieved after the time $T/2$, just before initiating the deceleration. From $\dot{x}(t) = gt$, we obtain

$$v_{\max} = g\frac{T}{2} = \sqrt{gD}.$$

Plugging in the different values for D results in

a) X = moon: $T \approx 3,5$ h, $v_{\max} \approx 63$ km/s.
b) X = Mars: $T \approx 42 - 112$ hours, $v_{\max} \approx 742 - 1980$ km/s.
c) X = Proxima Centauri: $T \approx 4$ years, $v_{\max} \approx 2,1\ c$.
d) X = Andromeda galaxy: $T \approx 2784$ years, $v_{\max} \approx 1434\ c$.

1.15. We perform the computations in the inertial frame of the earth. The earth's world line is then $\mathbb{R}\mathbf{e_0}$ and we choose the coordinates such that the destination X has the world line $\mathbb{R}\mathbf{e_0} + (0, D, 0, 0)$. Let $\tau \mapsto \mathbf{x}(\tau)$ the world line of the space ship, parametrized by proper time. The four-velocity is then

$$\mathbf{u}(\tau) = (\cosh(\varphi(\tau)), \sinh(\varphi(\tau)), 0, 0)$$

for a function $\varphi(\tau)$ yet to be determined. For the four-acceleration, we have

$$\mathbf{a}(\tau) = \varphi'(\tau)(\sinh(\varphi(\tau)), \cosh(\varphi(\tau)), 0, 0).$$

The absolute value of the four-acceleration is

$$g^2 = \langle\!\langle \mathbf{a}(\tau_0), \mathbf{a}(\tau_0)\rangle\!\rangle = (\varphi'(t))^2 \cdot 1,$$

therefore $\varphi'(\tau) = \pm g$ and hence $\varphi(\tau) = \pm g\tau + \varphi_0$. From $\mathbf{e_0} = \mathbf{u}(0) = (\cosh(\varphi_0), \sinh(\varphi_0), 0, 0)$ we conclude $\varphi_0 = 0$ and therefore $\varphi(\tau) = \pm g\tau$.
During the first half of the travel, the spacecraft accelerates in direction X. Thus $\varphi'(\tau) > 0$, hence $\varphi(\tau) = g\tau$. We conclude $\mathbf{u}(\tau) = (\cosh(g\tau), \sinh(g\tau), 0, 0)$ and therefore

$$\mathbf{x}(\tau) = \frac{1}{g}(\sinh(g\tau), \cosh(g\tau), 0, 0) + \mathbf{x_0}.$$

From

$$(0,0,0,0) = \mathbf{x}(0) = \frac{1}{g}(0,1,0,0) + \mathbf{x_0}$$

we get $\mathbf{x_0} = -\frac{1}{g}\mathbf{e_1}$. Summarizing we have the proper-time parametrization of the world line of the spacecraft:

$$\mathbf{x}(\tau) = \frac{1}{g}(\sinh(g\tau), \cosh(g\tau) - 1, 0, 0).$$

For the time of travel T_{ship} from the viewpoint of the crew we obtain

$$\frac{D}{2} = x^1\left(\frac{T_{\text{ship}}}{2}\right) = \frac{1}{g}\left(\cosh\left(g\frac{T_{\text{ship}}}{2}\right) - 1\right)$$

and hence

$$T_{\text{ship}} = \frac{2}{g}\operatorname{arcosh}\left(\frac{Dg}{2} + 1\right).$$

For the time of travel T_{earth} from the viewpoint of the earth we get

$$\begin{aligned}
\frac{T_{\text{earth}}}{2} &= x^0\left(\frac{T_{\text{ship}}}{2}\right) \\
&= x^0\left(\frac{1}{g}\operatorname{arcosh}\left(\frac{Dg}{2} + 1\right)\right) \\
&= \frac{1}{g}\sinh\left(\operatorname{arcosh}\left(\frac{Dg}{2} + 1\right)\right) \\
&= \frac{1}{g}\sqrt{\frac{g^2D^2}{4} + gD} \\
&= \sqrt{\frac{D^2}{4} + \frac{D}{g}}
\end{aligned}$$

hence

$$T_{\text{earth}} = \sqrt{D^2 + \frac{4D}{g}}.$$

Furthermore, the maximal velocity v_{max} can be calculated by

$$\begin{aligned}
v_{\text{max}} &= \frac{u^1(T_{\text{ship}}/2)}{u^0(T_{\text{ship}}/2)} \\
&= \tanh\left(\operatorname{arcosh}\left(\frac{gD}{2} + 1\right)\right) \\
&= \frac{\sqrt{\frac{g^2D^2}{4} + gD}}{\frac{gD}{2} + 1} \\
&= \frac{\sqrt{g^2D^2 + 4gD}}{gD + 2}.
\end{aligned}$$

Here, we always calculated in terms of the dimensionless mathematical velocity v, which is related to the physical velocity by the speed of light,

$$v = \frac{v_{\text{phys}}}{c}.$$

Mathematical length and time therefore have to have the same dimension. We choose the convention to calculate in units of length, i.e.

$$D = D_{\text{phys}}, \quad T = c \cdot T_{\text{phys}}, \quad g = \frac{g_{\text{phys}}}{c^2}.$$

Then we get

$$\begin{aligned}
T_{\text{ship, phys}} &= \frac{2c}{g_{\text{phys}}} \operatorname{arcosh}\left(\frac{D g_{\text{phys}}}{2c^2} + 1\right) \\
T_{\text{earth, phys}} &= \sqrt{\frac{D^2}{c^2} + \frac{4D}{g_{\text{phys}}}} \\
v_{\text{max, phys}} &= \frac{c\sqrt{D^2 {g_{\text{phys}}}^2 + 4D g_{\text{phys}} c^2}}{D g_{\text{phys}} + 2c^2}
\end{aligned}$$

Inserting the values for our destinations we get the following table:

		classical		*relativistic*		
object X	distance D	time T	v_{max}	T_{earth}	T_{ship}	v_{max}
moon	400.000 km	3,5 h	63 km/s	3,5 h	3,5 h	63 km/s
Mars (near)	56 mill. km	42 h	742 km/s	42 h	42 h	741 km/s
Mars (far)	400 mill. km	112 h	1980 km/s	112 h	112 h	1980 km/s
Proxima Centauri	4,3 light years	4 years	2,1 c	5,9 years	3,6 years	0,95 c
Andromeda galaxy	2 mill. light years	2784 years	1434 c	2 mill. years	28 years	almost c

B. SageMath computations

B.1. The Schwarzschild solution

B.1.1. The ansatz

Declaration of a mass parameter m and a 4-dimensional Lorentzian manifold M:

```
sage: reset()
```

None

```
sage: var('m', domain='real')
```

$$m$$

```
sage: assume(m>0)
```

None

```
sage: M = Manifold(4, 'M', structure='Lorentzian')
sage: std.<t,r,th,ph> = M.chart(r't r:(2*m,+oo) th:[0,pi]:\theta ph:[0,2*pi]:\phi')
```

Ansatz for the Schwarzschild metric:

```
sage: F = function('F')(r)
sage: G = function('G')(r)
sage: g = M.metric(name='g')
sage: g[0,0] = -F*F
sage: g[1,1] = 1
sage: g[2,2] = G*G
sage: g[3,3] = G*G*sin(th)^2
sage: g.display()
```

$$g = -F(r)^2\, \mathrm{d}t \otimes \mathrm{d}t + \mathrm{d}r \otimes \mathrm{d}r + G(r)^2\, \mathrm{d}\theta \otimes \mathrm{d}\theta + G(r)^2 \sin(\theta)^2\, \mathrm{d}\phi \otimes \mathrm{d}\phi$$

Compute the Christoffel symbols:

```
sage: g.christoffel_symbols_display()
```

$$\begin{array}{lcl}
\Gamma^{t}{}_{t\,r} & = & \frac{\frac{\partial F}{\partial r}}{F(r)} \\
\Gamma^{r}{}_{t\,t} & = & F\left(r\right) \frac{\partial F}{\partial r} \\
\Gamma^{r}{}_{\theta\,\theta} & = & -G\left(r\right) \frac{\partial G}{\partial r} \\
\Gamma^{r}{}_{\phi\,\phi} & = & -G\left(r\right) \sin\left(\theta\right)^{2} \frac{\partial G}{\partial r} \\
\Gamma^{\theta}{}_{r\,\theta} & = & \frac{\frac{\partial G}{\partial r}}{G(r)} \\
\Gamma^{\theta}{}_{\phi\,\phi} & = & -\cos\left(\theta\right) \sin\left(\theta\right) \\
\Gamma^{\phi}{}_{r\,\phi} & = & \frac{\frac{\partial G}{\partial r}}{G(r)} \\
\Gamma^{\phi}{}_{\theta\,\phi} & = & \frac{\cos(\theta)}{\sin(\theta)}
\end{array}$$

Show non-zero components of the Ricci tensor:

```
sage: g.ricci()[0,0]
```
15

$$\frac{F\left(r\right) G\left(r\right) \frac{\partial^{2} F}{\partial r^{2}} + 2\, F\left(r\right) \frac{\partial F}{\partial r} \frac{\partial G}{\partial r}}{G\left(r\right)}$$

```
sage: g.ricci()[1,1]
```
16

$$-\frac{G\left(r\right) \frac{\partial^{2} F}{\partial r^{2}} + 2\, F\left(r\right) \frac{\partial^{2} G}{\partial r^{2}}}{F\left(r\right) G\left(r\right)}$$

```
sage: g.ricci()[2,2]
```
17

$$-\frac{G\left(r\right) \frac{\partial F}{\partial r} \frac{\partial G}{\partial r} + F\left(r\right) \frac{\partial G}{\partial r}^{2} + F\left(r\right) G\left(r\right) \frac{\partial^{2} G}{\partial r^{2}} - F\left(r\right)}{F\left(r\right)}$$

```
sage: g.ricci()[3,3]
```
18

$$-\frac{\left(G\left(r\right) \frac{\partial F}{\partial r} \frac{\partial G}{\partial r} + F\left(r\right) \frac{\partial G}{\partial r}^{2} + F\left(r\right) G\left(r\right) \frac{\partial^{2} G}{\partial r^{2}} - F\left(r\right)\right) \sin\left(\theta\right)^{2}}{F\left(r\right)}$$

B.1.2. The Schwarzschild metric

Declaration of a mass parameter m and a 4-dimensional Lorentzian manifold M:

```
sage: reset()
```
19

None

```
sage: var('m', domain='real')
```
20

$$m$$

```
sage: assume(m>0)
```

None

```
sage: M = Manifold(4, 'M', structure='Lorentzian')
sage: std.<t,r,th,ph> = M.chart(r't␣r:(2*m,+oo)␣th:[0,pi]:\
   theta␣ph:[0,2*pi):\phi')
```

The Schwarzschild metric:

```
sage: def h(r):
....:     return 1-2*m/r
sage: g = M.metric(name='g')
sage: g[0,0] = -h(r)
sage: g[1,1] = 1/h(r)
sage: g[2,2] = r^2
sage: g[3,3] = r^2*sin(th)^2
sage: g.display()
```

$$g = \left(\frac{2\,m}{r} - 1\right) \mathrm{d}t \otimes \mathrm{d}t + \left(-\frac{1}{\frac{2\,m}{r} - 1}\right) \mathrm{d}r \otimes \mathrm{d}r + r^2 \mathrm{d}\theta \otimes \mathrm{d}\theta + r^2 \sin(\theta)^2 \,\mathrm{d}\phi \otimes \mathrm{d}\phi$$

The Christoffel symbols:

```
sage: g.christoffel_symbols_display()
```

$$\begin{array}{lcl}
\Gamma^{t}{}_{t\,r} & = & -\frac{m}{2\,mr - r^2} \\
\Gamma^{r}{}_{t\,t} & = & -\frac{2\,m^2 - mr}{r^3} \\
\Gamma^{r}{}_{r\,r} & = & \frac{m}{2\,mr - r^2} \\
\Gamma^{r}{}_{\theta\,\theta} & = & 2\,m - r \\
\Gamma^{r}{}_{\phi\,\phi} & = & (2\,m - r)\sin(\theta)^2 \\
\Gamma^{\theta}{}_{r\,\theta} & = & \frac{1}{r} \\
\Gamma^{\theta}{}_{\phi\,\phi} & = & -\cos(\theta)\sin(\theta) \\
\Gamma^{\phi}{}_{r\,\phi} & = & \frac{1}{r} \\
\Gamma^{\phi}{}_{\theta\,\phi} & = & \frac{\cos(\theta)}{\sin(\theta)}
\end{array}$$

Check that Schwarzschild is indeed Ricci flat:

```
sage: g.ricci().display()
```

$$\mathrm{Ric}\,(g) = 0$$

The Kretschmann scalar curvature:

```
sage: R = g.riemann()
sage: Km = R.down(g)['_{ijkl}'] * R.up(g)['^{ijkl}']
sage: Km.display()
```

$$\begin{array}{llll} M & \longrightarrow & \mathbb{R} \\ (t, r, \theta, \phi) & \longmapsto & \frac{48\,m^2}{r^6} \end{array}$$

B.2. The Kerr solution

B.2.1. Massless Kerr is Minkowski

We declare Minkowski spacetime:

```
sage: reset()
```

None

```
sage: Mink = Manifold(4, 'Mink', latex_name=r'\mathsf{Mink}')
sage: cart.<t,x,y,z> = Mink.chart()
sage: g = Mink.metric('g')
sage: g[0,0] = -1
sage: g[1,1] = 1
sage: g[2,2] = 1
sage: g[3,3] = 1
sage: g.display()
```

$$g = -\mathrm{d}t \otimes \mathrm{d}t + \mathrm{d}x \otimes \mathrm{d}x + \mathrm{d}y \otimes \mathrm{d}y + \mathrm{d}z \otimes \mathrm{d}z$$

We introduce weird polar coordinates which include angular momentum:

```
sage: var('a', domain='real')
```

$$a$$

```
sage: weird.<tt,r,theta,phi> = Mink.chart(r'tt␣r:(0,+oo)␣theta
    :(0,pi):\theta␣phi:(0,2*pi):\phi')
sage: trafo_weird_to_cart = weird.transition_map(cart, [tt,
    sqrt(r^2+a^2)*sin(theta)*cos(phi), sqrt(r^2+a^2)*sin(theta)
    *sin(phi), r*cos(theta)])
```

```
sage: trafo_weird_to_cart.display()                                      49
```

$$\begin{cases} t & = & tt \\ x & = & \sqrt{a^2 + r^2}\cos(\phi)\sin(\theta) \\ y & = & \sqrt{a^2 + r^2}\sin(\phi)\sin(\theta) \\ z & = & r\cos(\theta) \end{cases}$$

Now we check how the Minkowski metric looks in these coordinates:

```
sage: g.display(weird)                                                   50
```

$$g = -\mathrm{d}tt \otimes \mathrm{d}tt + \left(\frac{a^2\cos(\theta)^2 + r^2}{a^2 + r^2}\right)\mathrm{d}r \otimes \mathrm{d}r + \left(a^2\cos(\theta)^2 + r^2\right)\mathrm{d}\theta \otimes \mathrm{d}\theta + \left(a^2 + r^2\right)\sin(\theta)^2\,\mathrm{d}\phi \otimes \mathrm{d}\phi$$

This is precisely the Kerr metric with mass $m = 0$ and angular momentum a.

B.2.2. Check the Boyer-Lindquist identities

Declaration of the Kerr spacetime:

```
sage: reset()                                                            51
```

None

```
sage: var('m␣a', domain='real')                                          52
```

$$(m, a)$$

```
sage: Kerr = Manifold(4, 'Kerr', structure='Lorentzian')                 53
```

We introduce the Boyer-Lindquist coordinates:

```
sage: BL.<t,r,theta,phi> = Kerr.chart(r't␣r␣theta:(0,pi):\               54
    vartheta␣phi:(-oo,oo):\varphi')
sage: BLF = Kerr.default_frame()                                         55
```

We define the metric in Boyer-Lindquist coordinates:

```
sage: var('Delta␣rho')                                                   56
```

$$(\Delta, \rho)$$

```
sage: rho2 = r^2 + a^2*cos(theta)^2                                      57
sage: delta = r^2 - 2*m*r + a^2                                          58
sage: g = Kerr.metric('g')                                               59
sage: g[0,0] = -1+2*m*r/rho2                                             60
```

```
sage: g[0,3] = -2*m*r*a*sin(theta)^2/rho2
sage: g[1,1] = rho2/delta
sage: g[2,2] = rho2
sage: g[3,3] = (r^2+a^2+2*m*r*a^2*sin(theta)^2/rho2)*sin(theta
   )^2
sage: g[:]
```

$$\begin{pmatrix} \frac{2\,mr}{a^2\cos(\vartheta)^2+r^2}-1 & 0 & 0 & -\frac{2\,amr\sin(\vartheta)^2}{a^2\cos(\vartheta)^2+r^2} \\ 0 & \frac{a^2\cos(\vartheta)^2+r^2}{a^2-2\,mr+r^2} & 0 & 0 \\ 0 & 0 & a^2\cos(\vartheta)^2+r^2 & 0 \\ -\frac{2\,amr\sin(\vartheta)^2}{a^2\cos(\vartheta)^2+r^2} & 0 & 0 & \left(\frac{2\,a^2mr\sin(\vartheta)^2}{a^2\cos(\vartheta)^2+r^2}+a^2+r^2\right)\sin(\vartheta)^2 \end{pmatrix}$$

We check the Boyer-Lindquist identities (Lemma 4.8):

```
sage: g[3,3] + a*sin(theta)^2*g[0,3] - (r^2+a^2)*sin(theta)^2
```

$$0$$

```
sage: g[0,3] + a*sin(theta)^2*g[0,0] + a*sin(theta)^2
```

$$0$$

```
sage: a*g[3,3] + (r^2+a^2)*g[0,3] - delta*a*sin(theta)^2
```

$$0$$

```
sage: a*g[0,3] + (r^2+a^2)*g[0,0] + delta
```

$$0$$

Lemma 4.9 can also be checked directly:

```
sage: g[0,0]*g[3,3] - g[0,3]^2 + delta*sin(theta)^2
```

$$0$$

```
sage: det(g[:]) + rho2^2*sin(theta)^2
```

$$0$$

B.2.3. The Christoffel symbols

Show the nonzero Christoffel symbols. As we did for the Schwarzschild metric we could use the command `g.christoffel_symbols_display()` but we want to prevent SageMath from expanding all occurrances of Δ and ρ^2. For an expression this can be achieved by applying the method `subs(delta==Delta, rho2==rho^2)`. This is why we display the Christoffel symbols by hand using an iterated loop.

```
sage: nabla = g.connection()
sage: for i in Kerr.irange():
....:     for j in Kerr.irange():
....:         for k in range(j+1):
....:             result = factor(nabla[i,j,k].expr()).subs(
   delta==Delta, rho2==rho^2)
....:             if not result == 0:
....:                 show(r'$\Gamma^'+str(latex(BL[i]))+r'_{'
   +str(latex(BL[j]))+'␣'+str(latex(BL[k]))+r'}='+latex(result
   )+'$\n')
```

$$\Gamma^{t}_{rt} = \frac{\left(a^2 \sin(\vartheta)^2 - a^2 + r^2\right)\left(a^2 + r^2\right)m}{\Delta\rho^4}$$

$$\Gamma^{t}_{\vartheta t} = -\frac{2\,a^2 mr \cos(\vartheta)\sin(\vartheta)}{\rho^4}$$

$$\Gamma^{t}_{\varphi r} = \frac{\left(a^4 \cos(\vartheta)^2 - a^2 r^2 \cos(\vartheta)^2 - a^2 r^2 - 3\,r^4\right)am \sin(\vartheta)^2}{\Delta\rho^4}$$

$$\Gamma^{t}_{\varphi\vartheta} = -\frac{2\left(a^2 \sin(\vartheta)^2 - a^2 - r^2\right)a^3 mr \cos(\vartheta)\sin(\vartheta)^3}{\rho^6}$$

$$\Gamma^{r}_{tt} = -\frac{(a\cos(\vartheta)+r)(a\cos(\vartheta)-r)\Delta m}{\rho^6}$$

$$\Gamma^{r}_{rr} = \frac{a^2 m \cos(\vartheta)^2 - a^2 r \cos(\vartheta)^2 + a^2 r - mr^2}{\Delta\rho^2}$$

$$\Gamma^{r}_{\vartheta r} = -\frac{a^2 \cos(\vartheta)\sin(\vartheta)}{\rho^2}$$

$$\Gamma^{r}_{\vartheta\vartheta} = -\frac{\Delta r}{\rho^2}$$

$$\Gamma^{r}_{\varphi t} = \frac{(a\cos(\vartheta)+r)(a\cos(\vartheta)-r)\Delta am \sin(\vartheta)^2}{\rho^6}$$

$$\Gamma^{r}_{\varphi\varphi} = -\frac{\left(a^4 r\cos(\vartheta)^4 + a^4 m \cos(\vartheta)^2 \sin(\vartheta)^2 + 2\,a^2 r^3 \cos(\vartheta)^2 - a^2 mr^2 \sin(\vartheta)^2 + r^5\right)\Delta \sin(\vartheta)^2}{\rho^6}$$

$$\Gamma^{\vartheta}_{tt} = -\frac{2\,a^2 mr \cos(\vartheta)\sin(\vartheta)}{\rho^6}$$

$$\Gamma^{\vartheta}_{rr} = \frac{a^2 \cos(\vartheta)\sin(\vartheta)}{\Delta\rho^2}$$

$$\Gamma^{\vartheta}_{\vartheta r} = \frac{r}{\rho^2}$$

$$\Gamma^{\vartheta}_{\vartheta\vartheta} = -\frac{a^2 \cos(\vartheta)\sin(\vartheta)}{\rho^2}$$

$$\Gamma^{\vartheta}_{\varphi t} = \frac{2\left(a^2 + r^2\right)amr \cos(\vartheta)\sin(\vartheta)}{\rho^6}$$

$$\Gamma^{\vartheta}_{\varphi\varphi} = -\frac{\left(a^6 \cos(\vartheta)^4 - 2\,a^4 mr \cos(\vartheta)^4 + a^4 r^2 \cos(\vartheta)^4 + 2\,a^4 r^2 \cos(\vartheta)^2 - 4\,a^2 mr^3 \cos(\vartheta)^2 + 2\,a^2 r^4 \cos(\vartheta)^2 + 2\,a^4 mr + 4\,a^2 mr^3 + a^2 r^4 + r^6\right)\cos(\vartheta)\sin(\vartheta)}{\rho^6}$$

$$\Gamma^{\varphi}_{rt} = -\frac{(a\cos(\vartheta)+r)(a\cos(\vartheta)-r)am}{\Delta\rho^4}$$

$$\Gamma^{\varphi}_{\vartheta t} = -\frac{2\,amr\cos(\vartheta)}{\rho^4 \sin(\vartheta)}$$

$$\Gamma^{\varphi}_{\varphi r} = -\frac{a^4 m \cos(\vartheta)^4 - a^4 r \cos(\vartheta)^4 - a^4 m \cos(\vartheta)^2 + a^2 m r^2 \cos(\vartheta)^2 - 2\,a^2 r^3 \cos(\vartheta)^2 + a^2 m r^2 + 2\,m r^4 - r^5}{\Delta \rho^4}$$

$$\Gamma^{\varphi}_{\varphi \vartheta} = \frac{\left(a^4 \sin(\vartheta)^4 - 2\,a^4 \sin(\vartheta)^2 + 2\,a^2 m r \sin(\vartheta)^2 - 2\,a^2 r^2 \sin(\vartheta)^2 + a^4 + 2\,a^2 r^2 + r^4\right)\cos(\vartheta)}{\rho^4 \sin(\vartheta)}$$

B.2.4. The Killing tensor field

Definition of the canonical vector field V:

```
sage: V = Kerr.vector_field(name='V')
sage: V[0] = r^2 + a^2
sage: V[3] = a
sage: V.display()
```

$$V = \left(a^2 + r^2\right)\frac{\partial}{\partial t} + a\frac{\partial}{\partial \varphi}$$

Two lightlike vector fields:

```
sage: Lplus = 1/delta*V + BLF[1]
sage: Lplus.display()
```

$$\left(\frac{a^2 + r^2}{a^2 - 2\,mr + r^2}\right)\frac{\partial}{\partial t} + \frac{\partial}{\partial r} + \left(\frac{a}{a^2 - 2\,mr + r^2}\right)\frac{\partial}{\partial \varphi}$$

```
sage: Lminus = 1/delta*V - BLF[1]
sage: Lminus.display()
```

$$\left(\frac{a^2 + r^2}{a^2 - 2\,mr + r^2}\right)\frac{\partial}{\partial t} - \frac{\partial}{\partial r} + \left(\frac{a}{a^2 - 2\,mr + r^2}\right)\frac{\partial}{\partial \varphi}$$

Now we define the symmetric $(0,2)$-tensor field K:

```
sage: K = delta/2*(Lplus.down(g)*Lminus.down(g)+Lminus.down(g)
    *Lplus.down(g)) + r^2*g
sage: K.display()
```

$$\left(\frac{a^2 r^2 + \left(a^4 - 2\,a^2 mr\right)\cos(\vartheta)^2}{a^2\cos(\vartheta)^2 + r^2}\right)\mathrm{d}t\otimes\mathrm{d}t + \left(-\frac{2\,a^3 mr \sin(\vartheta)^4 - \left(2\,a^3 mr - a^3 r^2 - ar^4 - \left(a^5 + a^3 r^2\right)\cos(\vartheta)^2\right)\sin(\vartheta}{a^2\cos(\vartheta)^2 + r^2}$$

Compute its covariant differential:

```
sage: nablaK = nabla(K)
sage: nablaK.display()
```

$$ar\sin(\vartheta)^2\,\mathrm{d}t\otimes\mathrm{d}r\otimes\mathrm{d}\varphi + \left(a^3 - 2\,amr + ar^2\right)\cos(\vartheta)\sin(\vartheta)\,\mathrm{d}t\otimes\mathrm{d}\vartheta\otimes\mathrm{d}\varphi - ar\sin(\vartheta)^2\,\mathrm{d}t\otimes\mathrm{d}\varphi\otimes\mathrm{d}r - \left(a^3 - 2\,amr + ar^2\right)$$

Now we define a generic vector field and check that K is a Killing tensor field:

```
sage: var('T R TH PH')                                          91
```

$$(T, R, TH, PH)$$

```
sage: X = Kerr.vector_field(name='X')                           92
sage: X[0] = T                                                  93
sage: X[1] = R                                                  94
sage: X[2] = TH                                                 95
sage: X[3] = PH                                                 96
sage: X.display()                                               97
```

$$X = T\frac{\partial}{\partial t} + R\frac{\partial}{\partial r} + TH\frac{\partial}{\partial \vartheta} + PH\frac{\partial}{\partial \varphi}$$

```
sage: nablaK(X,X,X).display()                                   98
```

$$\begin{array}{llll} Kerr & \longrightarrow & \mathbb{R} \\ (t, r, \vartheta, \varphi) & \longmapsto & 0 \end{array}$$

B.2.5. The Carter constant

We find two expressions for the Carter constant. The first one is

```
sage: q = g(X,X).expr()                                         99
sage: v = T*delta - PH*delta*a*sin(theta)^2                     100
sage: Carter1 = r^2*q - rho2^2*R^2/delta + v^2/delta            101
sage: Carter1                                                   102
```

$$-\frac{\left(a^2 \cos(\vartheta)^2 + r^2\right)^2 R^2}{a^2 - 2\,mr + r^2} - \frac{\left(2\,TH^2 mr^5 - TH^2 r^6 - 2\,T^2 a^2 mr - 4\,T^2 mr^3 - \left(TH^2 a^2 + R^2 - T^2\right) r^4 - \left(TH^2 a^6 - 2\,TH^2 a^4 mr + TH^2\right.}{}$$

```
sage: Error1 = K(X,X).expr() - Carter1                          103
sage: Error1.simplify_full()                                    104
```

$$0$$

The second one is

```
sage: w = -T*a*sin(theta)^2 + PH*(r^2+a^2)*sin(theta)^2         105
sage: Carter2 = rho2^2*TH^2 + w^2/sin(theta)^2 - a^2*q*cos(     106
   theta)^2
sage: Carter2                                                   107
```

$$\left(a^2 \cos(\vartheta)^2 + r^2\right)^2 TH^2 + \frac{\left(2\,TH^2 mr^5 - TH^2 r^6 - 2\,T^2 a^2 mr - 4\,T^2 mr^3 - \left(TH^2 a^2 + R^2 - T^2\right) r^4 - \left(TH^2 a^6 - 2\,TH^2 a^4 mr + TH^2\right.}{}$$

```
sage: Error2 = K(X,X).expr() - Carter2
sage: Error2.simplify_full()
```

$$0$$

B.2.6. Ricci curvature

Check that the Ricci curvature of Kerr vanishes, i.e. that Kerr is a vacuum solution:

```
sage: g.ricci().display()
```

$$\mathrm{Ric}\,(g) = 0$$

Bibliography

[1] C. Bär: *Elementargeometrie*, Lecture Notes, University of Potsdam, Potsdam 2008, https://www.math.uni-potsdam.de/fileadmin/user_upload/Prof-Geometrie/Dokumente/Lehre/Lehrmaterialien/skript-ElemGeo.pdf

[2] C. Bär: *Differential Geometry*, Lecture Notes, University of Potsdam, Potsdam 2006, https://www.math.uni-potsdam.de/fileadmin/user_upload/Prof-Geometrie/Dokumente/Lehre/Lehrmaterialien/skript-DiffGeo-engl.pdf

[3] A. Einstein: *Zur Elektrodynamik bewegter Körper*, Ann. der Phys. (4) 17, 891–921 (1905), https://doi.org/10.1002/andp.19053221004

[4] J.C. Hafele and R. E. Keating: *Around-the-World Atomic Clocks: Predicted Relativistic Time Gains*, Science 177, issue 4044, 166–168 (1972), https://doi.org/10.1126/science.177.4044.166

[5] J.C. Hafele and R. E. Keating: *Around-the-World Atomic Clocks: Observed Relativistic Time Gains*, Science 177, issue 4044, 168–170 (1972), https://doi.org/10.1126/science.177.4044.168

[6] C. W. Misner, K. S. Thorne and J. A. Wheeler: *Gravitation*, W. H. Freeman and Company, New York (1973)

[7] B. O'Neill: *Semi-Riemannian Geometry - With Applications to Relativity*, Academic Press, San Diego (1983)

[8] B. O'Neill: *The Geometry of Kerr Black Holes*, Dover Publ., Mineola (1995)

[9] H. Stephani: *Relativity. An Introduction to Special and General Relativity*, 3rd Edition, Cambridge University Press, Cambridge (2004)

[10] R. M. Wald: *General relativity*, The University of Chicago Press, Chicago-London (1984)

Index

Symbols

A

I

K

L

M

N

O