# General Relativity

## for Differential Geometers

### with emphasis on world lines rather than space slices

## Philadelphia, Spring 2007


## Hermann Karcher, Bonn


**Contents**

## Preface

When the Mathematics Department of the University of Pennsylvania contacted me to spend a term with them, I discussed with Chris Croke and Wolfgang Ziller plans for a course topic. They thought that a course on Relativity, addressed to graduate students in differential geometry, would find most interest. This turned out to be the case and the interest I met encouraged me to write these notes. The notes, while written as a differential geometric text, do develop many applications until observable numbers are obtained.

For the preparation of this course I had substantial help from the summaries that my former students wrote for each of my lectures in the summer 1994. I hope that these extended Philadelphia notes will find their way back to some of them. Their enthusiasm motivated me to suggest to Chris and Wolfgang that such a course might work again. – The advice I got from Jürgen Ehlers, Peter Schneider and Andreas Quirrenbach was essential for my background in Astrophysics, i.e., for the words to be said between the mathematics.

There are nonessential differences to other expositions and one essential one. The fact that I wrote for an audience with a good working knowledge in differential geometry is irrelevant for the contents, adaption to other audiences is therefore straight forward. However, there is one deviation from other texts which is more than a matter of taste: I have heavily emphasized world lines and de-emphasized space slices. The reason is that our clocks are now good enough to measure proper time on the clocks' world lines with enough precision to show relativistic effects. And they are also precise enough so that definitions of rest spaces of observers, definitions that go back to Einstein in Special Relativity, do not work in less linear situations, e.g. in the Schwarzschild geometry.

I am grateful to my colleagues at Penn and to the graduate students I met for creating such a friendly and interested atmosphere in which it was a pleasure to work. In addition, many thanks to Herman Gluck for all the help in other matters.

Philadelphia, Spring 2007
Hermann Karcher

# Einstein's Clocks

How can identical clocks measure time at different rates?

Einstein's theory of Special Relativity started with thought experiments that analyzed the concept of simultaneity. It took 50 years before more and more experiments started to be performed that verified Einstein's predictions to higher and higher accuracy. 25 years ago relativity entered daily life when the global positioning system (GPS) was built. But many people still react with complete disbelief to the statement that time may pass with different speed – nowadays measurably in many situations. I will first try to explain what the precise meaning of this statement is and then show that this basic fact of relativity theory is in perfect agreement with other facts from physics that are less difficult to accept.

## What are clocks?

The first precision time pieces were pendulum clocks. They had one imperfection that caused problems in astronomy and made them useless on the ocean: When transported they lost their precision completely. Time pieces with balance springs were much better behaved and quartz clocks essentially did not loose precision when transported. These classical clocks have a common principle: They have a very regular but delicate clock pulse generator, a mechanism that counts the pulses, translates the count into and shows the time that passed, and finally an energy source that keeps the pulse generator going. Presently our standard time is measured with atomic clocks. Basically the clock pulse generator is the transition frequency between two energy levels of the element cesium and the point is that the transition radiation has an extraordinarily narrow band width. More technically, a microwave radiation of approximately this transition frequency is synthesized and its absorption by the cesium atoms is used to regulate it to precisely the correct transition frequency. Again, the frequency is counted and the count is translated into time that passed. Let me now emphasize that it does not really matter what opinions about time one has. But one needs to realize that all statements involving time in physics mean the time that is **measured**, presently by cesium clocks. For example, the physicists and the engineers involved in the installation of the global positioning system did not agree about how time passes. Therefore two different counting systems had to be installed in the early GPS satellites. The non-relativistic version was so far off that the system did not work.

The choice of the element cesium for our standard clocks has technical reasons for achieving high precision. In principle one can use the transition frequency between any two energy levels of any atom. I connect this fact with a fundamental astronomical observation: If one observes spectral lines in the light of any celestial object, then one can identify subfamilies of lines as the lines of specific elements **because the ratios of the celestial lines are the same as the ratios in our laboratories!** Rephrased as time measurements this says: The atomic clocks at any place in the universe (that we have been able to observe) agree among each other about how time passes at that place. The fact that the ratios agree and not the frequencies themselves means that we are observing clocks which agree among each other but their time passes with different speed than ours. We will see that relative motion, the so called Doppler effect, can explain this.

Observation of identical clocks that tick with different speed.

What kind of an experiment could one imagine that lets us observe identical clocks ticking with different speed? In principal we could sit next to one clock and observe another one ticking differently. A skeptic would still blame the clock rather than accept our statement about time passing differently. Therefore I want to describe another type of clock that, admittedly, cannot be built to the precision of a cesium clock, but they convey such a robust impression of the passing of time that I find it very difficult to disbelieve them. These clocks measure the passing of time with radioactive material: one unit of time of such a clock has passed, if one half of the original amount of material has decayed. Several such clocks are in practical use in archaeology. There is no indication so far that they might not agree with the cesium clocks. Now, if we hand to two physicists equal blocks of radium, let them go their ways and when they later meet again we count the radium atoms they have left. If one of them has 5% fewer than the other aren't we forced to say that for him 5% more time has passed? – Well, except for the skeptical remark: I would prefer to see such an experiment instead of speculating about its possible outcome.

Already when I was a student the physicists Pound and Rebka performed such an experiment. They put one (generalized) atomic clock on the ground floor of a 40 m tower and an identical clock at the top. The bottom clock sent its time signals to the top. Technically simpler, the bottom clock sent directly the transition radiation of its clock pulse generator to the top clock. The newly discovered Mösbauer effect had to be used so that the emitted radiation did not loose momentum to the emitting atom. At the top Pound and Rebka observed that the incoming frequency was too slow to be absorbed by the identical atoms of the top clock pulse generator. In other words: they observed that the bottom clock was ticking more slowly than the top clock! Even more surprising, they could determine how much too slow the bottom clock was and found that the difference in clock speed was in perfect agreement with older well established facts from physics. Here are the details:
From an electromagnetic wave one can absorb energy only in portions

$$E = h \cdot \nu$$

These portions are called photons.
These photons have a mass $m$

$$E = m \cdot c^2,\ m = E/c^2 = h \cdot \nu / c^2$$

Here $c$ denotes the speed of light.
If some mass $m$ flies a distance $s$ upwards in the gravitational field of the earth, then it looses the following amount of kinetic energy

$$\Delta E = m \cdot g \cdot s.$$

Pound and Rebka found that also photons loose exactly this much energy! We can translate this energy change into a frequency change:

$$\Delta\nu = \frac{\Delta E}{h} = \nu \cdot \frac{g \cdot s}{c^2},$$

and it is exactly this frequency change that was responsible for the bottom clock to tick more slowly. (Summary on transparency 1 at the end.)

> In other words, if we accept the two Nobel prize formulas above and the Pound Rebka measurement (made possible by the Mösbauer Nobel prized effect) then the time signals of the bottom clock arrive at the top clock at the slowed down rate predicted by energy conservation!

Clocks in motion relative to each other.

Now we turn to the origin of Special Relativity. Decades before cesium clocks and the Pound Rebka experiment Einstein predicted on the basis of thought experiments that relative motion would affect how identical clocks measure time. History shows that many people are unable to accept Einstein's analysis (among them were even the engineers of the GPS project). I believe one reason is that we have absolutely no every day experience with observations made by two people in fairly fast motion relative to each other. Therefore I chose the Pound Rebka experiment as introduction: An observer can sit quietly and watch the two identical clocks tick at different rate. This situation is so simple that one cannot argue with its set up.

In Einstein's 1905 analysis there was no gravity. We are asked to imagine two observing physicists in whose laboratories one cannot measure the faintest traces of any acceleration. However they are allowed to be in constant relative motion. As far as we know, the laws of physics have to be exactly the same in all such situations. This is now postulated as **the principle of relativity** and neither experiments nor theoretical analysis raise any suspicion that this principle might be wrong. Such laboratories are called inertial systems. Note that such inertial systems are an idealization which does not exist in our world. Einstein's falling elevator can only turn off a strictly homogenous gravitational field, not the real fields we live in. Therefore no practical reference frame will be strictly inertial. Special Relativity is part of the ideal world of inertial systems and its acceptance has to be in this idealized form. Its assumptions are never strictly satisfied, in no real or imagined laboratory of our world.

Let me recall that the situation is the same with our 3-dimensional Euclidean geometry. We are completely at ease in using this ideal geometry, although we can never check whether our physical surroundings strictly satisfy its axioms. Let me recall one property of Euclidean geometry which is very similar to what we will meet in Special Relativity. We are accustomed to use coordinates called Height, Width and Depth, they measure distances in three orthogonal directions. Given these orthogonal measurements we compute the length $\ell$ of a vector $(x, y, z)$ with the Pythagorean theorem as $\ell = \sqrt{x^2 + y^2 + z^2}$. Then we discover that this formula is not tied to our standard coordinates: we can take any three pairwise orthogonal unit vectors $\{e_1, e_2, e_3\}$, write $(x, y, z) = x_1e_1 + x_2e_2 + x_3e_3$ and find the surprising fact that the length is always computed by the same formula: $\ell = \sqrt{x_1^2 + x_2^2 + x_3^2}$. In other words, although we usually think of having a naturally preferred coordinate system it is true that all the other coordinates are equally good and no geometric difference between them exists. In a completely analogous way we will describe the geometry of Special Relativity first from the point of view of one preferred inertial observer and then we discover that in all inertial systems the same formulas hold. The analogy goes still farther. Of course we know from our Euclidean geometry the following: If we join two distinct points in space by two different curves then we find it silly to expect that the two curves have the same length. If we accept the geometry of Special Relativity with the same trust then the famous twin paradox goes away by turning silly: the time measured by a clock is the length of that curve that describes the traveling life of the clock, and length means *length with respect to the geometry of Special Relativity.* As in the Euclidean analogue: it is silly to expect that different curves have the same length.

To derive the geometry of Special Relativity we only use the **principle of relativity** and a fundamental hypothesis formulated by Einstein: *The traveling speed of a light signal is independent of the motion of its source*, or in more colloquial words: the **speed of light is constant**. Physicists had met this constant traveling speed of electromagnetic waves already before Einstein, in Maxwell's theory of electromagnetism. And briefly before Einstein published 'Special Relativity', further support was given to the constant speed of light hypothesis by the (negative) result of the Michelson-Morley interferometer experiment.

## The Geometry of Special Relativity.

What we have to understand can be condensed into the following main problem. Consider two inertial observers whose inertial systems have the velocity $v$ relative to each other. We assume further that they meet at some moment and set their clocks to zero at that instant. When their clocks show time 1 each of them sends a light signal towards the other one (moving away with velocity $v$).

> The time $T$ when these light signals are received will be the same for both of these inertial observers because of the relativity principle.

| How large is $T$? |
|---|

To answer this question they agree to return a light signal at the moment when the first signal is received (i.e. at clock time $T$). The first signals were sent at clock time 1 and received at clock time $T$. For the second pair of signals the time intervals are stretched by a factor $T$: sent at clock time $T$ and received at clock Time $T^2$. Both of them use the same clocks, hence the same units of time. To measure lengths they agree on units such that the speed of light is $c = 1$. Now both of them can plot the world line of the other and the world lines of the light signals in coordinates of their inertial systems, see transparency 2 at the end. Our observers solve two linear equations and find $T^2$, hence $T$, in terms of $c$ and $v$.

This fundamental relation gives the factor $T$ by which the time between two received light signals is longer than the time between the emissions of these signals (positive $v$ means moving apart). This frequency shift is called the **Doppler effect**.

$$T^2 = \frac{c+v}{c-v}.$$

Now both can mark on the world line of the other(!) the points where the clock time is one. They now observe that for **all inertial observers** the Time-1-Points on the other world lines satisfy (in their own coordinates) the equation of a two-sheeted hyperboloid:

$$t^2 - |x|^2 = 1.$$

The two physicists therefore have achieved for Special Relativity what corresponds, in our Euclidean 3-space, to the determination of the unit ball.
The quadratic expression $t^2 - |x|^2$ plays for Special Relativity the same role that the Pythagorean theorem plays for Euclidean space. In particular it determines the time-like arc length on world lines without reference to any(!) observer. But this time-like arc length on a world line is the time that an atomic clock having this world line does measure: *Measured time is a geometric property of the world line in question.*

In the last statement we apply the insight that we obtained for inertial observers more generally to accelerated observers, in other words: to curved world lines. We justify this generalization by noting that the time-like arc length of a curved world line can be obtained by approximating the world line by piecewise straight, i.e. non-accelerated, world lines. Since the corners of such approximations are not physically meaningful one might also want to see experimental support. Indeed, we can observe particles with a very short lifetime circling at high speed in a synchroton. Not only do we notice immediately that they circle many more times than their lifetime permits, we also find after doing the computation (see transparency 3 at the end) that the number of completed orbits is exactly what the computed passing of time on these world lines allows them. Notice that this is a twin paradox experiment: a twin particle watching from the center of the synchroton its orbiting twin will reach the end of its life time long before the orbiting particle decays. Put differently, it is not difficult to imagine two physicists starting their rather different lives with two equal chunks of radium. When they meet again late in life it would be a colossal coincidence if the time-like arc lengths of their world lines really were the same. Therefore they will find their remaining chunks of radium to be of different size.

One can even observe the different passing of time in (fairly) inertial systems and on inertial world lines. Of course, in such a situation the two world lines cannot have the same start point and the same end point. For a full explanation it would therefore be necessary to discuss how distances are measured in the two inertial systems. This requires more definitions than just clocks. Therefore we only mention the experiment without detailed explanation:

Collisions by cosmic rays generate high in the atmosphere very short lived but fast traveling mesons. They are measured in a laboratory about 30 kilometers away. Even with the speed of light they could not travel 30 km in their life time. However the Time-1-Point for their world line is given by Minkowski's hyperbola and the result is that much less proper time passes on the meson's world line from the top of the atmosphere to the ground laboratory than passes on the world line of a rocket that flies between the same places. Therefore its life time suffices to reach the ground.

Summary and repetition:
1.) Since, according to Pound and Rebka, photons flying upwards in a gravitational field loose the same amount of (kinetic) energy as a mass $m = h\nu/c^2$ gains in potential energy, the frequency $\nu$ of the corresponding wave is decreased by the same percentage. This can be rephrased by saying: the distance between time signals increases by the same percentage. Therefore we watch the clock that is higher up in the gravitational field ticking faster by exactly this percentage.
2.) The principle of relativity and the constancy of the speed of light imply that the Time-1-Points on unaccelerated world lines in inertial systems lie on the gauge surface $t^2 - |x|^2 = 1$. This allows to define a time-like arc length on world lines and our analysis of clocks means that this time-like arc length is the (so called proper) time that passes along such a world line and is measured by atomic clocks or decaying radium.

Experiments that support Special Relativity:
    http://www.atomki.hu/fizmind/specrel/experiments.html
Clock debate before the start of GPS satellites:
    http://www.leapsecond.com/history/Ashby-Relativity.htm

Transparency 1

**1.) According to Max Planck one can absorb energy from an electromagnetic wave only in portions**

$$E = h \cdot \nu.$$

**These portions are called photons.**

**2.) These photons have a mass** $m$ **according to Einstein's famous formula**

$$\textbf{in general: } E = m \cdot c^2, \quad \textbf{for photons: } m = \frac{h \cdot \nu}{c^2},$$

**where** $c$ **denotes the speed of light.**

**3.) If some mass** $m$ **flies a height** $s$ **upwards in the gravitational field of the earth, then it looses the following amount of kinetic energy**

$$\Delta E = m \cdot g \cdot s.$$

**4.) The experiment of Pound and Rebka shows that the same is true for photons**

$$\Delta E = \frac{h \cdot \nu}{c^2} \cdot g \cdot s.$$

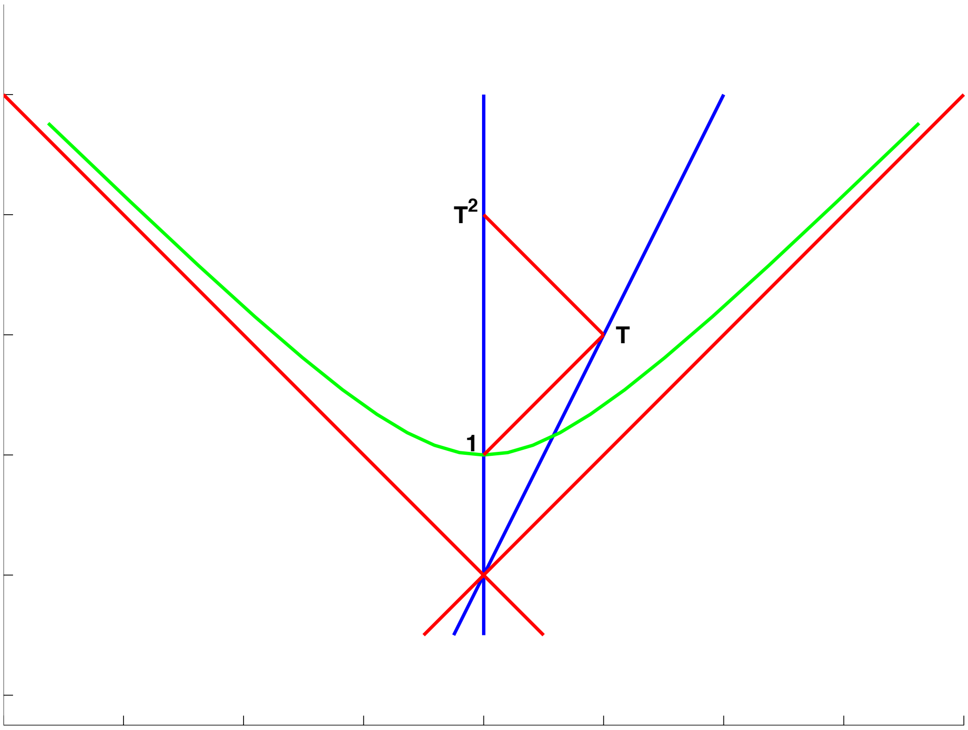
**This energy change translates into a frequency change**

$$\Delta \nu = \frac{\Delta E}{h} = \nu \cdot \frac{g \cdot s}{c^2}.$$

**5.) A clock which is a height** $s$ **above another clock in the field of the earth ticks faster by this same percentage**

$$\frac{\Delta \nu}{\nu} = \frac{g \cdot s}{c^2}!$$

**The Time 1 Points of Minkowski Geometry**




**1.) World line of a light signal starting from 1 (red):** $\binom{0}{1} + \binom{1}{1} \cdot t$
**The world line of the second observer, starting at 0 (blue):** $\binom{a}{1} \cdot s$
**The intersection of these two world lines (at yet unknown clock time $T$) is:**

$$\begin{pmatrix}0\\1\end{pmatrix} + \begin{pmatrix}1\\1\end{pmatrix} \cdot \frac{a}{1-a} = \begin{pmatrix}a\\1\end{pmatrix} \cdot \frac{1}{1-a}.$$

**2.) The returning signal is received at clock time $T^2$ in**

$$\begin{pmatrix}0\\T^2\end{pmatrix} = \begin{pmatrix}0\\\frac{1+a}{1-a}\end{pmatrix}, \textbf{hence } T = \sqrt{\frac{1+a}{1-a}}.$$

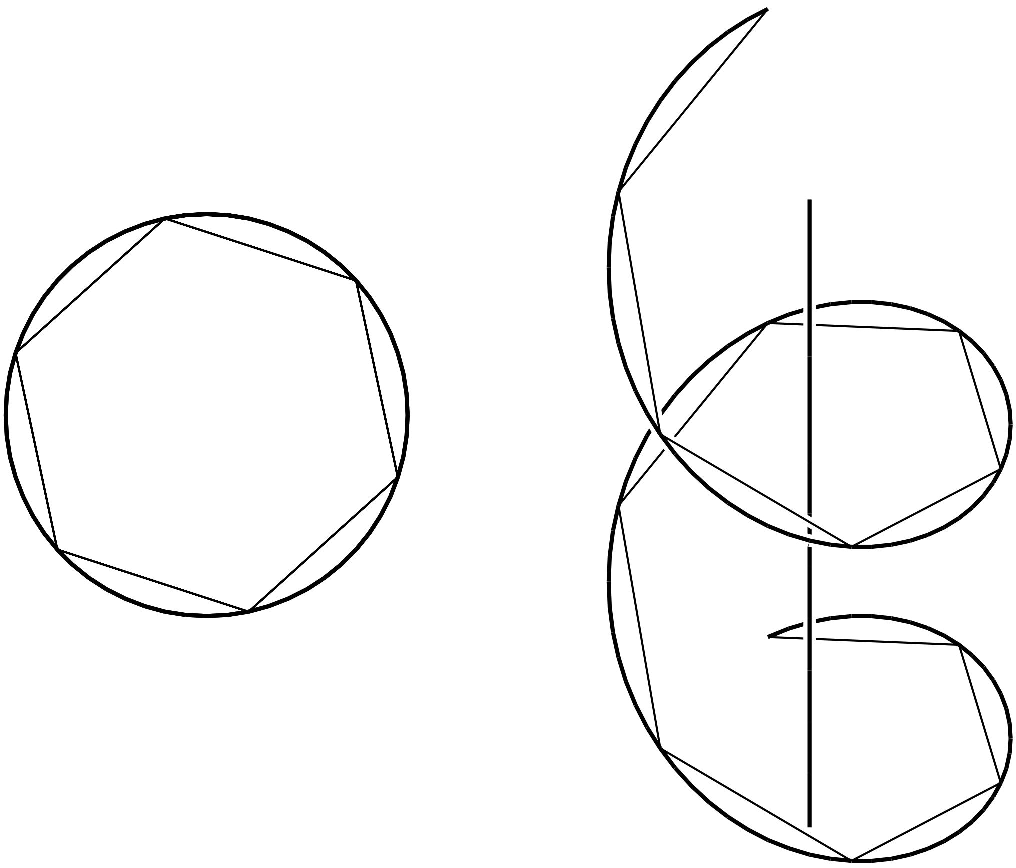
**3.) The Time-1-Point on the second world line therefore is at**

$$\frac{1}{T} \cdot \begin{pmatrix}a\\1\end{pmatrix} \cdot \frac{1}{1-a} = \frac{1}{\sqrt{1-a^2}} \cdot \begin{pmatrix}a\\1\end{pmatrix}.$$

**4.) All the Time-1-Points $\binom{x}{t}$ (green) therefore satisfy the following hyperbola equation**

$$t^2 - x^2 = 1.$$

**How Time passes in a Synchroton**



**It is a theorem that the arc length of smooth curves in Euclidean space can be determined via approximation by polygons. The same proof shows that the time-like arc length of world lines can be determined via approximation by piecewise non-accelerated worldlines, even though the corners of these approximations are physically unrealistic. Since we have found the Time-1-Points on straight world lines we can conclude how time passes on the world lines of particles circling in the synchroton. Such a world line is a**

$$\textbf{helix: } c(s) := \begin{pmatrix} \cos(s) \\ \sin(s) \\ h \cdot s \end{pmatrix}, \qquad h > 1$$

**and time passes as**

$$T(s) = \sqrt{h^2 - 1} \cdot s,$$

**while on the world line that is the axis of the helix the larger time** $T_{axis} = h \cdot s$ **passes.**

**It is a correct idea to imagine time as time-like arc length of world lines.**

## Indefinite Scalar Products

Isometry Groups, Geodesics on Spheres, Space Time Coordinates

For the interpretation of the machinery of Relativity some intuitive understanding of indefinite scalar products is required.

In terms of the standard scalar product $\langle .,. \rangle$ on $\mathbb{R}^n$ we can define indefinite scalar products with the help of a diagonal matrix $S$

$$S := \text{diag } (+1, ..p\text{-times}.., +1, -1, ..(n-p)\text{-times}.., -1) \quad \text{as}$$

$$\langle\langle u, v \rangle\rangle := \langle S \cdot u, v \rangle \quad \text{or} \quad \langle\langle u, v \rangle\rangle = \sum_{i=1}^{p} u_i v_i - \sum_{i=p+1}^{n} u_i v_i.$$

Linear maps $A : \mathbb{R}^n \to \mathbb{R}^n$ are called *isometries* or $\langle\langle .,. \rangle\rangle$-*orthogonal* if they satisfy

$$u, v \in \mathbb{R}^n \Rightarrow \langle\langle Au, Av \rangle\rangle = \langle\langle u, v \rangle\rangle.$$

We define quadratic surfaces (called *generalized spheres*)

$$Q_\pm := \{v \in \mathbb{R}^n : \langle\langle v, v \rangle\rangle = \pm 1\}.$$

Of course we cannot avoid to look at pictures with eyes trained in Euclidean geometry. Therefore one should note the following

Transitivity Theorem

The isometries of $(\mathbb{R}^n, \langle\langle .,. \rangle\rangle)$ are transitive on $Q_+$ and on $Q_-$.

Proof. Writing $\langle\langle u, v \rangle\rangle = \sum_{i=1}^{p} u_i v_i - \sum_{i=p+1}^{n} u_i v_i$ we have $O(p) \times O(n-p)$ as that subgroup of the isometry group which a Euclidean trained eye can observe immediately. For $x := (x_1, \dots, x_p, x_{p+1}, \dots, x_n) \in Q_\pm$ we put $A^2 := \sum_{i=1}^{p} x_i^2, \quad B^2 := \sum_{i=p+1}^{n} x_i^2$. Because of the transitivity of the groups $O(p), O(n-p)$ on the spheres $\mathbb{S}^{p-1}$ resp. $\mathbb{S}^{n-p-1}$ we can isometrically move $x$ to $(A, 0, \dots, 0, B)$. Since $x \in Q_\pm$ means $A^2 - B^2 = \pm 1$we can write $A, B$ as $\cosh\tau, \sinh\tau$. It remains to show that $(\cosh\tau, 0, \dots, 0, \sinh\tau) \in Q_+$ can isometrically be moved to $(1, 0, \dots, 0) \in Q_+$ and $(\sinh\tau, 0, \dots, 0, \cosh\tau) \in Q_-$ can isometrically be moved to $(0, \dots, 0, -1)$. Indeed, the matrix

$$M := \begin{pmatrix} \cosh\tau & 0 & \dots & 0 & -\sinh\tau \\ 0 & 1 & & & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & & & 1 & 0 \\ \sinh\tau & 0 & \dots & 0 & -\cosh\tau \end{pmatrix}$$

does this. And $M$ is an isometry because

$$x_1^2 - x_n^2 = (\cosh\tau x_1 - \sinh\tau x_n)^2 - (\sinh\tau x_1 - \cosh\tau x_n)^2.$$

Note that it is enough to check squares since also for indefinite scalar products we have the polarization identity:

$$4\langle\langle u, v \rangle\rangle = \langle\langle u + v, u + v \rangle\rangle - \langle\langle u - v, u - v \rangle\rangle.$$

Tangent Space Restriction Theorem

The restriction of the bilinear form $\langle\langle ., . \rangle\rangle$ from $\mathbb{R}^n$ to a tangent space of $Q_+$ resp. $Q_-$ looses a +sign (i.e. has signature $(p-1, n-p)$) on $Q_+$ respectively looses a $-$sign (i.e. has signature $(p, n-p-1)$) on $Q_-$.

Proof. Because of the transitivity theorem we only need to check this on the tangent space at $(1, 0, \ldots, 0) \in Q_+$ resp. $(0, \ldots, 0, 1) \in Q_-$, where it is trivial.

Definition. The cone $LC := \{u \in \mathbb{R}^n : \langle\langle u, u \rangle\rangle = 0\}$ is called the *light cone* of the indefinite scalar product $\langle\langle ., . \rangle\rangle$.

Tangent Space Intersection Theorem

The intersection of $Q_\pm$ with one of its tangent spaces $T_pQ_\pm$ ist the light cone of the restriction of $\langle\langle ., . \rangle\rangle$ to this tangent space.

Proof. A curve $c(t) \subset Q_\pm$ through $p$ satisfies $\langle\langle c(t), c(t) \rangle\rangle = \pm 1,\ c(0) = p$, hence

$$0 = \frac{d}{dt} \langle\langle c(t), c(t) \rangle\rangle|_{t=0} = 2\langle\langle p, \dot{c}(0) \rangle\rangle.$$

We therefore have for the tangent vector space $T_pQ_\pm = \{v \in \mathbb{R}^n : \langle\langle p, v \rangle\rangle = 0\}$. The affine tangent space in $\mathbb{R}^n$ is $p + T_pQ_\pm$. Any point $u$ in the affine tangent space satisfies:

$$u = p + v, \langle\langle p, p \rangle\rangle = \pm 1,\ \langle\langle p, v \rangle\rangle = 0.$$

Therefore $u \in Q_\pm$, i.e. $\langle\langle u, u \rangle\rangle = \pm 1$, is equivalent with

$$0 = \langle\langle u, u \rangle\rangle - \langle\langle p, p \rangle\rangle = \langle\langle p, v \rangle\rangle + \langle\langle v, v \rangle\rangle = \langle\langle v, v \rangle\rangle,$$

which says that $u$ is in the light cone of the affine tangent space (with the restricted scalar product) or in other words that $v$ is in the light cone of the tangent vector space.

Oblique Reflection Fact

The map $M$ that was used in the transitivity proof clearly satisfies $M^2 = 1$. It obviously has $n-2$ Eigenvalues $+1$ (with the eigenvectors being vectors of the basis). The remaining eigenvalues are $+1$ and $-1$ with eigenvectors $(x_1, x_n)_+ = (\cosh \tau/2, \sinh \tau/2)$ and $(x_1, x_n)_- = (\sinh \tau/2, \cosh \tau/2)$. All eigenvectors together are an $\langle\langle ., . \rangle\rangle$-orthonormal basis. – Of course the last two eigenvectors are not orthogonal for the Euclidean metric that we naturally use for picture interpretations. But, as in the Euclidean case, $(\cosh \tau/2, \sinh \tau/2)$ is a point on the "indefinite sphere" $\{(x_1, x_n) : x_1^2 - x_n^2 = 1\}$ and the other eigenvector is a tangent vector to this sphere at that point.

If we use instead of the standard basis $\{e_1, \ldots, e_n\}$ the eigenbasis $\{e_+, e_2, \ldots, e_{n-1}, e_-\}$ in which $x \in \mathbb{R}^n$ has the coordinates $\{x_+, x_2, \ldots, x_{n-1}, +x_-\}$ then $Mx$ has the coordinates $\{x_+, x_2, \ldots, x_{n-1}, -x_-\}$ and $\langle\langle x, x \rangle\rangle = \langle\langle Mx, Mx \rangle\rangle$ is obvious.

Next we turn to geodesics on these quadratic surfaces. Because of the indefiniteness of the metric it does not make good sense to look for *shortest* curves. I will introduce the covariant derivative of an indefinite Riemannian metric soon. Presently we use the

Definition of Straightest Curves. A curve $c(t)$ on $Q_\pm$ (later: on a submanifold) is called a *straightest curve* or a *geodesic* if its acceleration $\ddot{c}(t)$ has no tangential component.

Planar Geodesics Theorem

The geodesics on $Q_\pm$ are – as in the case of the Euclidean sphere – intersection of $Q_\pm$ with 2-planes that pass through the origin of $\mathbb{R}^n$.

Proof. We abbreviate $\sigma := \pm 1$. For any curve $t \mapsto \gamma(t) \in Q_\sigma$ we have

$$\langle\langle \gamma(t), \gamma(t) \rangle\rangle = \sigma, \quad \text{hence} \quad \langle\langle \gamma(t), \dot{\gamma}(t) \rangle\rangle = 0, \quad \text{or} \quad \dot{\gamma}(t) \in T_{\gamma(t)} Q_\sigma.$$

Geodesics have by definition no tangential acceleration. Presently the tangent spaces are orthogonal to the position vector, hence $\ddot{\gamma}(t) = \lambda(t)\gamma(t)$. This implies

$$\frac{d}{dt}\langle\langle \dot{\gamma}(t), \dot{\gamma}(t) \rangle\rangle = 2\langle\langle \dot{\gamma}(t), \ddot{\gamma}(t) \rangle\rangle = 0, \quad \text{hence} \quad \langle\langle \dot{\gamma}(t), \dot{\gamma}(t) \rangle\rangle = \langle\langle \dot{\gamma}(0), \dot{\gamma}(0) \rangle\rangle$$

Similarly, differentiating $\langle\langle \gamma(t), \dot{\gamma}(t) \rangle\rangle = 0$ gives

$$0 = \langle\langle \dot{\gamma}(t), \dot{\gamma}(t) \rangle\rangle + \langle\langle \gamma(t), \ddot{\gamma}(t) \rangle\rangle = \langle\langle \dot{\gamma}(t), \dot{\gamma}(t) \rangle\rangle + \langle\langle \gamma(t), \lambda(t)\gamma(t) \rangle\rangle$$

and therefore we have $\lambda(t) = -\langle\langle \dot{\gamma}(0), \dot{\gamma}(0) \rangle\rangle / \sigma =: -k$. So we have obtained a simple differential equation for $\gamma(t)$:

$$\ddot{\gamma}(t) + k \cdot \gamma(t) = 0.$$

The solution can be written as linear combination of the initial conditions with the help of either trigonometric or hyperbolic functions. For unification I use the following

Function Definition: Denote the function that solves $\ddot{f} + k \cdot f = 0$ with initial conditions

$$f(0) = 1,\ f'(0) = 0 \quad \text{by} \quad \mathbf{c}_k(t)$$
$$f(0) = 0,\ f'(0) = 1 \quad \text{by} \quad \mathbf{s}_k(t).$$

General geodesic: $\gamma(t) := \gamma(0) \cdot \mathbf{c}_k(t) + \dot{\gamma}(0) \cdot \mathbf{s}_k(t)$. Clearly this curve is in the vector subspace spanned by $\{\gamma(0), \dot{\gamma}(0)\}$. QED

The quadratic surfaces $Q_\pm$ are analogous to spheres because of their definition in terms of the scalar product. They also share a curvature property with the Euclidean spheres. Recall the

Definition of the Weingarten Map (or Shape Operator): Let $N(.)$ be a unit normal field along a hypersurface. For each curve $t \mapsto c(t)$ in the hypersurface we put

$$S_{c(t)}\dot{c}(t) := \frac{d}{dt} N(c(t)) \perp N(c(t)).$$

Apply this definition using that for all $x \in Q_\pm$ the normal is $N(x) = x$. All eigenvalues of the shape operator of the hypersurface $Q_\pm$ are therefore 1 (this property is called *umbillic*):

$$S_{c(t)}\dot{c}(t) = \dot{c}(t) \quad \text{or} \quad S = \mathrm{id} : T_x Q_\pm \to T_x Q_\pm.$$

Light Cone Determines Metric Theorem

Two indefinite quadratic forms with the same light cone are proportional.

Proof. Observe that two quadratic forms that agree on an open set are equal. Choose a fixed timelike vector $v$, i.e. $\langle\langle v, v\rangle\rangle < 0$. For all $w$ from the open set of spacelike vectors, i.e. $\langle\langle w, w\rangle\rangle > 0$, consider the straight line $u(t) := v + tw$. For small $|t|$ the vectors $u(t)$ are timelike, for large $|t|$ they are spacelike. Each such line therefore hits the light cone twice. Let $q(.,.)$ be the other quadratic form, with the same light cone. Choose $\lambda(v)$ such that $q(v,v) - \lambda\langle\langle v, v\rangle\rangle = 0$. Define $b(.,.) := q(.,.) - \lambda\langle\langle .,.\rangle\rangle$ and observe that $t \mapsto b(v+tw, v+tw)$ is a quadratic polynomial with **three** zeros, one at $t = 0$, the other two on the light cone. These polynomials are therefore zero, in other words $b(u, u) = 0$ for an open set of $u$, hence $b = 0, q = \lambda \cdot \langle\langle .,.\rangle\rangle$. QED

Here are two reasons why we will meet conformal changes of the metric extensively: (i) The Maxwell equations, which control electromagnetic waves, are conformally invariant. (ii) The important cosmological models by Friedman are conformally flat. To get used to conformal changes we prove that stereographic projection is a conformal map.

Definition of Stereographic Projection. Let $p \in Q_\pm$ and let $T_pQ_\pm$ be the affine tangent space, i.e. we write its elements as $p + v$ with $\langle\langle p, v\rangle\rangle = 0$. The stereographic projection projects $p + v$ from the point $-p \in Q_\pm$ (opposite to $p$) to $Q_\pm$. In other words, the line

$$g(t) := -p(1-t) + (p+v)t = -p + (2p+v)t$$

intersects $Q_\pm$ in $-p = g(0)$ and in $St(v)$, at $t = 4\sigma/(4\sigma + \langle\langle v, v\rangle\rangle)$.

Stereographic Projection:

$$St(v) := -p + (2p+v)\frac{4\sigma}{(4\sigma + \langle\langle v, v\rangle\rangle)}$$

Conformality Theorem: Stereographic projection is conformal.

Proof. The statement means that every derivative is a linear conformal map. We expand $St(v + \Delta v) = St(v) + Lin(\Delta v) + O\big((\Delta v)^2\big)$ and have to prove that the linear term is conformal.

$$\begin{aligned} St(v+\Delta v) &= -p + (2p + v + \Delta v)\frac{4\sigma}{4\sigma + \langle\langle v+\Delta v, v+\Delta v\rangle\rangle} \\ &= St(v) + \frac{4\sigma}{4\sigma + \langle\langle v, v\rangle\rangle}\Big(\Delta v - 2\frac{(2p+v)\langle\langle v, \Delta v\rangle\rangle}{4\sigma + \langle\langle v, v\rangle\rangle}\Big) + O\big((\Delta v)^2\big) \end{aligned}$$

Next observe that $4\sigma + \langle\langle v, v\rangle\rangle = \langle\langle 2p+v, 2p+v\rangle\rangle$ and abbreviate $x := 2p + v$. Then

$$St(v+\Delta v) - St(v) = \frac{4\sigma}{4\sigma + \langle\langle v, v\rangle\rangle}\Big(\Delta v - 2\frac{x\langle\langle x, \Delta v\rangle\rangle}{\langle\langle x, x\rangle\rangle}\Big) + O\big((\Delta v)^2\big),$$

where the linear term $\Delta v \to \Big(\Delta v - 2\frac{x\langle\langle x, \Delta v\rangle\rangle}{\langle\langle x, x\rangle\rangle}\Big)$ is an isometry, hence conformal. QED

$$SL(2,\mathbb{R}) := \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} : ad - bc = 1 \right\}.$$

We rename $a, b, c, d$ as

$$a := x_1 + x_4, \quad b := x_2 + x_3, \quad c := x_3 - x_2, \quad d := x_1 - x_4, \quad \text{hence}$$
$$1 = ad - bc = x_1^2 + x_2^2 - x_3^2 - x_4^2 = \langle\langle x, x \rangle\rangle.$$

The following curves (=subgroups) are geodesics – check $\ddot{\gamma}(t) + k\gamma(t) = 0$:

$$t \mapsto \begin{pmatrix} e^t & 0 \\ 0 & e^{-t} \end{pmatrix} \qquad t \mapsto \begin{pmatrix} \cosh t & \sinh t \\ \sinh t & \cosh t \end{pmatrix} \qquad t \mapsto \begin{pmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{pmatrix}.$$

In all cases $\gamma(0) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \text{id}$, or $x = (1, 0, 0, 0)$, and

$$\dot{\gamma}(0) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \qquad \dot{\gamma}(0) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad \dot{\gamma}(0) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},$$
$$= (0, 0, 0, 1) \qquad\qquad = (0, 0, 1, 0) \qquad\qquad = (0, 1, 0, 0).$$

All geodesics through the identity are 1-parameter subgroups. But not all points are reached by these geodesics: Because they are subgroups we have $\gamma(t) = \gamma(t/2) \cdot \gamma(t/2)$ so that every reached point is a square – but while $\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$, only slightly different ones are not squares: $\begin{pmatrix} -r & 0 \\ 0 & -1/r \end{pmatrix}$, $r > 1$. In other words: *This very simple and nice example does not have the Hopf-Rinow property*! This indicates that geodesic completeness (which is such a natural assumption in Riemannian geometry) may NOT be so useful in the indefinite cases. Indeed, except for the Minkowski space of Special Relativity, all our astronomically interesting examples are not complete.

Our characterization of geodesics as 2-plane sections of $Q_\pm$ allows to discuss completeness further: Any pair of non-antipodal points $p, q \in Q_\pm$ determines exactly one 2-plane and it cuts out the *only* geodesic through $p$ and $q$. If this geodesic is a hyperbola and $p, q$ are on different components then they cannot be joined by a geodesic.

Next we look at different parametrizations of $Q_+ \subset \mathbb{R}^5$ for the scalar product

$$\langle\langle x, y \rangle\rangle := \sum_{i=1}^{4} x_i y_i - x_5 y_5.$$

The restriction of this scalar product to the tangent spaces of $Q_+$ has the signature of Special Relativity. Our parametrizations of $Q_+$ can be described as different families of timelike geodesics. The distribution given as the $\langle\langle .,. \rangle\rangle$-orthogonal completement of the tangent vectors of these geodesics is integrable. We therefore get a foliation of $Q_+$ by "spaces" that are quite different.

1. Parametrization:

$$Q_+ = \left\{ \begin{pmatrix} e \cdot \cosh \tau \\ \sinh \tau \end{pmatrix} : e \in \mathbb{S}^3, \tau \in \mathbb{R} \right\} \text{ and the curves } \tau \mapsto \begin{pmatrix} e \cdot \cosh \tau \\ \sinh \tau \end{pmatrix}$$

are the timelike geodesics mentioned above. Now we compute the induced scalar product using this parametrization. For a curve $\gamma(.)$ we have

$$\gamma(t) := \begin{pmatrix} e(t) \cdot \cosh \tau(t) \\ \sinh \tau(t) \end{pmatrix}, \quad \dot{\gamma}(t) = \begin{pmatrix} \dot{e}(t) \cdot \cosh \tau \\ 0 \end{pmatrix} + \begin{pmatrix} e(t) \cdot \sinh \tau(t) \\ \cosh \tau(t) \end{pmatrix} \cdot \dot{\tau}(t)$$

and hence (using $e(t) \perp \dot{e}(t)$)

$$\langle\langle \dot{\gamma}(t), \dot{\gamma}(t) \rangle\rangle = -\dot{\tau}(t)^2 + \cosh^2 \tau \cdot \langle \dot{e}(t), \dot{e}(t) \rangle_{\mathbb{S}^3}.$$

The spacelike time slices $\{\tau = const\} = \{x_5 = const\}$ are 3-dimensional round spheres with equator length $2\pi \cosh \tau$.

2. Parametrization:

$$Q_+ \supset \left\{ q(\omega, \rho, \tau) := \begin{pmatrix} \omega \sinh \rho \sinh \tau \\ \cosh \tau \\ \cosh \rho \sinh \tau \end{pmatrix} : \tau \in \mathbb{R}, \omega \in \mathbb{S}^2, \rho \in \mathbb{R}_+ \right\}$$

and the curves $\tau \mapsto q(\omega, \rho, \tau)$ are the timelike geodesics mentioned above. Now we compute the induced scalar product using this parametrization. For any curve $\gamma(t) := q(\omega(t), \rho(t), \tau(t))$ we have

$$\dot{\gamma}(t) = \begin{pmatrix} \dot{\omega}(t) \cdot \sinh \rho \sinh \tau \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} \omega \cosh \rho \sinh \tau \\ 0 \\ \sinh \rho \sinh \tau \end{pmatrix} \cdot \dot{\rho}(t) + \begin{pmatrix} \omega \sinh \rho \cosh \tau \\ \sinh \tau \\ \cosh \rho \cosh \tau \end{pmatrix} \cdot \dot{\tau}(t),$$

hence (using $\omega \perp \dot{\omega}$)

$$\langle\langle \dot{\gamma}(t), \dot{\gamma}(t) \rangle\rangle = -\dot{\tau}(t)^2 + \sinh^2 \tau \cdot (\dot{\rho}^2 + \sinh^2 \rho \langle \dot{\omega}(t), \dot{\omega}(t) \rangle_{\mathbb{S}^2}).$$

Here the time slices $\{\tau = const\} = \{x_4 = const\}$ are 3-dimensional hyperbolic spaces of curvature $-1/\sinh^2 \tau$. – Note that these coordinates do not cover all of $Q_+$ although the hyperbolic spaces and the timelike geodesics are complete, except for the singularity at $\tau = 0$.

3. Parametrization:

$$Q_+ \supset \left\{ q(\omega, u, \tau) := \begin{pmatrix} \omega u \exp \tau \\ \cosh \tau - 0.5u^2 \exp \tau \\ \sinh \tau + 0.5u^2 \exp \tau \end{pmatrix} : \tau \in \mathbb{R}, \omega \in \mathbb{S}^2, u \in \mathbb{R}_+ \right\}$$

and the curves $\tau \mapsto q(\omega, u, \tau)$ are the timelike geodesics mentioned above. Now we compute the induced scalar product using this parametrization. For any curve $\gamma(t) := q(\omega(t), u(t), \tau(t))$ we have

$$\dot\gamma(t) = \begin{pmatrix} \dot\omega(t) \cdot u \exp\tau \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} \omega \exp\tau \\ -u \exp\tau \\ +u \exp\tau \end{pmatrix} \cdot \dot u(t) + \begin{pmatrix} \omega u \exp\tau \\ \sinh\tau - 0.5u^2 \exp\tau \\ \cosh\tau + 0.5u^2 \exp\tau \end{pmatrix} \cdot \dot\tau(t),$$

hence (using $\omega \perp \dot\omega$ and the orthogonality of the last two vectors)

$$\langle\langle \dot\gamma(t), \dot\gamma(t) \rangle\rangle = -\dot\tau(t)^2 + \exp 2\tau \cdot (\dot u^2 + u^2 \langle \dot\omega(t), \dot\omega(t) \rangle_{\mathbb{S}^2}).$$

Here the time slices $\{\tau = \mathit{const}\} = \{x_4 + x_5 = \mathit{const}\}$ are 3-dimensional Euclidean spaces, their metric given in polar coordinates $u, \omega$. Again, the coordinates do not cover all of $Q_+$, this time without showing a coordinate singularity, and as before with complete timelike geodesics and complete Euclidean spaces.

4. Parametrization:

$$Q_+ \supset \left\{ q(\omega, \alpha, \tau) := \begin{pmatrix} \omega \sin\alpha \\ \cos\alpha \cosh\tau \\ \cos\alpha \sinh\tau \end{pmatrix} : \alpha \in [0, \pi/2), \omega \in \mathbb{S}^2, \tau \in \mathbb{R} \right\}.$$

Here the timelike parameter lines $\tau \mapsto q(\omega, \alpha, \tau)$ are neither geodesics nor is $\tau$ the timelike arc length on them. We list this example because it appears as limit in a family of astronomically interesting examples. We compute the induced scalar product using this parametrization. For any curve $\gamma(t) := q(\omega(t), \alpha(t), \tau(t))$ we have

$$\dot\gamma(t) = \begin{pmatrix} \dot\omega \sin\alpha \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} \omega \cos\alpha \\ -\sin\alpha \cosh\tau \\ -\sin\alpha \sinh\tau \end{pmatrix} \cdot \dot\alpha(t) + \begin{pmatrix} 0 \\ \cos\alpha \sinh\tau \\ \cos\alpha \cosh\tau \end{pmatrix} \cdot \dot\tau(t),$$

hence, using the orthogonality of these vectors we get

$$\langle\langle \dot\gamma(t), \dot\gamma(t) \rangle\rangle = -\cos^2(\alpha)\dot\tau(t)^2 + \dot\alpha^2 + \sin^2(\alpha) \langle \dot\omega(t), \dot\omega(t) \rangle_{\mathbb{S}^2}.$$

The slices $\{\tau = \mathit{constant}\} = \{\sinh\tau \cdot x_4 = \cosh\tau \cdot x_5\}$ are unit 3-spheres parametrized in polar coordinates.

---

One can look at similar examples on $Q_- \subset \mathbb{R}^5$. Since we want the product on the tangent spaces to have the signature of Special Relativity one takes as product on $\mathbb{R}^5$

$$\langle\langle x, x \rangle\rangle := x_1^2 + x_2^2 + x_3^2 - x_4^2 - x_5^2.$$

## Special Relativity

Minkowski Diagrams, Simultaneity, Distance, Compton Effect, Center of Mass

We come back to the discussion at the end of lecture 1, but we explain the basic definitions in more detail. We consider first one preferred inertial observer. His *world* is $\mathbb{R}^4 = \mathbb{R}^3 \times \mathbb{R}$ and each *event* has coordinates $(x, y, z, t)$. The statement that his laboratory is inertial means that a mass point $(x, y, z)$ that

a) is at rest at time $t_0$ and

b) has no forces acting on it

stays at rest. We express this by saying

**The world line of a point at rest is** $t \mapsto (x, y, z, t)$.

Being an inertial system means a little more. If the mass point moves at initial time $t = 0$ with velocity $v = (v_1, v_2, v_3)$ then it continues to do so, i.e.

The world line of a force free mass point is $t \mapsto (x + v_1 t, y + v_2 t, z + v_3 t, t)$.

Today, time measurements are more precise than length measurements. Therefore the unit of time, the second, is defined first, namely in terms of the frequency of the cesium transition that is used in our standard clocks. The meter is defined as the distance which light travels in an agreed fraction of a second. For drawing diagrams it is best to take the unit of length such that the speed of light is 1. The world lines of light signals in the coordinate space of our preferred observer are therefore straight lines with slope 1.

Constant Speed of Light Hypothesis. **The world lines of light signals do not depend on the source that emitted the signals.** – In other words: when we look at the world line of a light signal we can draw no conclusion about the velocity of the emitting source.

Next we connect these statements with observations. The experiments will be described assuming the *Constant Speed of Light Hypothesis*. This hypothesis is therefore not checked directly. It is supported indirectly because all our predictions about the outcome of experiments agree with the measurements.

*How can two inertial observers check that they are relative to each other at rest?*

(i) Light signal travel times are constant: One observer sends a light signal to the other. The other observer returns the signal upon arrival. The first observer records the round trip travel time of the signal and checks whether this time is constant.

(ii) Angle sizes of known objects are constant: The second observer shows the first observer an object known to both of them. The first observer measures under which angle he sees the known object. (For example, the angle under which we see the sun is about one half of a degree.) This angle size has to remain constant in time.

(iii) Baseline measurements give constant distance: The first observer has two telescopes and he knows the distance between them. He points both telescopes to the same point that the second observer shows him. For each telescope he measures the angle between the direction in which this telescopes points and the direction to the other telescope. In other words, he determines a triangle from one edge and the two adjacent angles; the distance is the distance to the opposite (the third) corner of this triangle. This distance has to remain constant in time.

If the two inertial observers have constant velocity relative to each other then the above measurements give that the measured distance changes linearly in time. – However this is not the usual way in which relative velocity is measured. Physicists and Astronomers use the fact that the time between two light signals is different for the emitter and the receiver if the two have a velocity relative to each other. This very important phenomenon is called the *Doppler effect*. We derive its size with the help of a second basic assumption of Special Relativity:

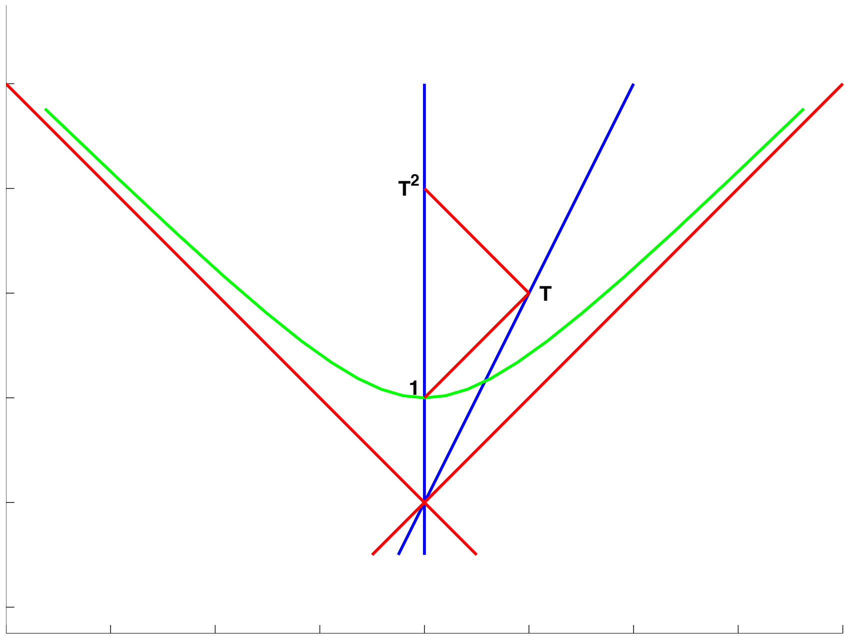
INERTIAL OBSERVER HYPOTHESIS, also referred to as RELATIVITY PRINCIPLE.

**The laws of physics are the same for any two inertial observers.**

Example of its application. If two inertial observers fly away from each other with velocity $v$ and each of them sends two light signals to the other that are one second apart then the time $T$ between the received signals *is the* **same** *for both observers.*

We now look once again at the discussion in the first lecture:

**Doppler Effect and Time 1 Points of Minkowski Geometry**




1.) World line of a light signal starting from 1 (red): $\binom{0}{1} + \binom{1}{1} \cdot t$

The world line of the second observer, starting at 0, relative velocity $v$ (blue): $\binom{v}{1} \cdot s$

The intersection of these two world lines (at yet unknown clock time $T$) is:

$$\begin{pmatrix}0\\1\end{pmatrix} + \begin{pmatrix}1\\1\end{pmatrix} \cdot \frac{v}{1-v} = \begin{pmatrix}v\\1\end{pmatrix} \cdot \frac{1}{1-v}.$$

2.) The returning signal is received at clock time $T^2$ in

$$\begin{pmatrix} 0 \\ T^2 \end{pmatrix} = \begin{pmatrix} 0 \\ \frac{1+v}{1-v} \end{pmatrix}, \textbf{hence Doppler Ratio: } T = \sqrt{\frac{1+v}{1-v}}.$$

3.) The Time-1-Point on the second world line therefore is at

$$\frac{1}{T} \cdot \begin{pmatrix} v \\ 1 \end{pmatrix} \cdot \frac{1}{1-v} = \frac{1}{\sqrt{1-v^2}} \cdot \begin{pmatrix} v \\ 1 \end{pmatrix}.$$

4.) All the **Time-1-Points** $\binom{x}{t}$ (green) therefore satisfy the following hyperbola equation

$$t^2 - x^2 = 1.$$

---

As a consequence of the two stated physical hypotheses we have obtained via a harmless computation two very important consequences. There is first the Doppler effect that can now be used to

MEASURE RELATIVE VELOCITY:

> If two observers fly away from each other ( $v > 0$ in this case) then the time difference between received light signals is larger by the factor $T = \sqrt{(1+v)/(1-v)}$ (*Doppler Ratio*, or *Doppler Red Shift*,) than the time difference between the emitted signals.

Secondly, the determination of the Time-1-Points on the world lines of all inertial observers determines the MINKOWSKI GEOMETRY of SPECIAL RELATIVITY. In the language of this geometry all inertial observers are treated in the same away, in particular, *there is no more distinction of the originally preferred observer!* The importance of having a geometric picture based on the most important notion – Time – of Relativity cannot be overestimated.

---

We have almost finished the discussion of the passage of time. Only a short addition is needed, an addition that is in perfect agreement with experiments with short lived particles that fly in fields that bend their world lines. Any curve $c(s) = (x(s), y(s), z(s), t(s))$ in our Minkowski geometry that has everywhere a slope larger than 1, or in other words

$$\langle\langle c'(s), c'(s)\rangle\rangle = x'(s)^2 + y'(s)^2 + z'(s)^2 - t'(s)^2 < 0,$$

can be the world line of a clock. The time measured by the clock is called **proper time**, it is the timelike arc length of the world line:

$$\textbf{Proper Time} = \int_{t_0}^{t_1} \sqrt{-\langle\langle c'(s), c'(s)\rangle\rangle} ds$$

Proper time is a geometric quantity that depends only on the world line of the measuring clock, no other oberservers or their clocks are involved. Already at this point there is no “twin paradox” left, only “twin facts”: The proper times along two world lines that have the same initial points and the same final points will be different from each other in most cases (and whether you notice this or not only depends on the precision of your clock).

**Synchronization of clocks relative at rest to each other**

Einstein started his discussion of Special Relativity with thought experiments that determine what we mean by: *synchronization of clocks which are at rest relative to each other.* My view of the history is that the reason for so many people to have difficulty in accepting Relativity Theory lies in this synchronization discussion. Einstein's proposal depends heavily on the constant speed of light hypothesis and there is nothing in our daily life experiences that has any similarity with this constancy of the speed of light. .

EINSTEIN'S THOUGHT EXPERIMENT. Assume we have two inertial observers relative at rest to each other in our Minkowski world. This means they have parallel straight lines as their world lines and the natural clocks on these world lines tick with the same speed (given by the same Time-1-Point on the hyperboloid $|x|^2 - t^2 = -1$). The question is:

*which points on the two world lines are simultaneous?*

Einstein's answer is simple (but only possible because of the constant speed of light):

| |
|---|
| Events $A_1$ on the first, $A_2$ on the second world line are *simultaneous*, if *light signals emitted at* $A_1, A_2$ *reach the middle world line simultaneously!* |

Since our model is the vector space $\mathbb{R}^4$ there is no ambiguity about what the middle world line is. (One can also invoke more light signals: (a) when looking from the first world line then the middle world line should be seen in the same direction as the second world line and (b) the light signal travel times from the middle world line to either the first or the second world line should be the same.)

Einstein's answer leads to a geometrically satisfying statement:
The set of events that are simultaneous with an event $A$ on a straight world line is the $\langle\langle ., . \rangle\rangle$-orthogonal complement through $A$ of the straight world line.
Moreover:
The light signal travel time distance between parallel world lines is the geometric length of the segment between simultaneous events on these two world lines. (Reflection in the simultaneity space of $(v, 1)$ is the map given by $(1, v) \mapsto (1, v)$, $(v, 1) \mapsto (-v, -1)$, it preserves the light cone.)

Note that the discussion of clock synchronization also led to a geometric statement about distances. For emphasis I repeat:
We only measure the distance between observers relative at rest to each other - well, of course, the two ends of the original prototype meter in Paris where at rest relative to each other. And, the light signal travel time distance between the world lines of two observers relative at rest has a geometric interpretation: Pick two *simultaneous* events $A_1, A_2$ on these two parallel world lines, then

$$\text{Light Signal Travel Time Distance between World Lines} =$$
$$= \sqrt{\langle\langle A_1 - A_2, A_1 - A_2 \rangle\rangle} \;=\; \text{Geometric Distance between } A_1, A_2.$$

With these explanations we are ready to finish the discussion from the first lecture concerning the short lived particles that cannot travel 30 km in their life time but that are nevertheless measured 30 km from where they are created.

Consider the two parallel world lines $(0, 0, 0, t)$ and $(30, 0, 0, t)$ of two, say, clocks that are at rest in the inertial system of the observing physicist and have a distance of 30 km from each other. A particle flying with relative velocity $v$ may have the world line $(v \cdot t, 0, 0, t)$ that meets the first one at $t = 0$ and the second one at $t = 30/v$. The proper time $s = \sqrt{1 - v^2} \cdot t$ on the particle's world line is clearly less than $t$ and quite a bit so if $v$ is close to 1. Now, the particle stays (statistically: 50%) alive as long as this proper time is shorter than its (half) life time, so that these particles can easily travel from one to the other world line. *I repeat:* the key to understand this experiment is to clearly realize that there is no "universal" passing of time, but time passes locally, for each world line as the geometry dictates.

One should also ask: *how much distance does the particle travel?* Recall that this distance is by definition the light signal travel time distance between the world line of the particle and another parallel world line through some event $B$ in the space that is simultaneous with the event $A = (0, 0, 0, 0)$ of the particle at the moment when it starts. What is $B$? We intersect the orthogonal complement of $(v, 0, 0, 1)$ with the world line $(30, 0, 0, t)$, more specifically, we intersect $(1, 0, 0, v) \cdot r$ with $(30, 0, 0, t)$, This gives $B = (30, 0, 0, t = 30v)$. It is important to note that $A$ and $B$ are indeed simultaneous for the particle, but $B$ is **later** than $A$ for the physicist. The geometric length of $B - A$ is $30\sqrt{1 - v^2}$, obviously considerably shorter than 30 km if $v$ is close to 1, but, a pleasant surprise, equal to $v \cdot s$. This fact is often translated into non-physics language by saying that a measuring rod that moves with velocity $v$ is shortened by the factor $\sqrt{1 - v^2}$. This completely disregards that this shortened length is measured in a different inertial system, a system in which the initial point of the rod and the end point of the rod are no longer simultaneous for the original observer. I have not seen a discussion how one should achieve this miracle: keep the initial point of the rod at the event $A$ and transport the end point from $(30, 0, 0, 0)$ into the future, to $B$. Already at this early stage of Relativity Theory the geometric language is so superior that such mysteries do not even arise.

TIME, VELOCITY, DISTANCE SUMMARY. Our primary measurements are time measurements. This includes frequency changes, when we use the Doppler effect to determine relative velocities (even the police does it this way). Finally, distances are defined and measured via light signal travel times. The original meter is now an object of only historic interest.

The isometries of the indefinite metric of Special Relativity are called **Lorentz Transformations**. They were known before Einstein from Maxwell's theory of electromagnetism. It was Minkowski who formulated the geometry of Special Relativity. Nevertheless it is called Lorentz Geometry because of Minkowsi's contributions in the geometry ofnumbers.

The main examples of relativistic mechanics are collision experiments with elementary particles. I will treat the *Compton effect*, the collision of a particle of mass $m_0$ that is at rest at the origin with a photon of energy $h\nu$ flying in the $x$-direction. After the collision the photon has energy $h\nu'$ and its direction of flight makes an angle $\alpha$ with the $x$-axis, the particle flies with velocity $v$ under an angle $\beta$. To formulate what equation has to be satisfied we need the notion of *energy-momentum vector*. For a particle with restmass $m_0 \neq 0$ we take the timelike unit vector of its world line and multiply it by $m_0$. For a photon we have to agree on the inertial system in which we want to give its energy as $h\nu$. In this coordinate system we multiply the nullvector $(1, 0, 0, 1)$ tangent to its world line with $h\nu$. Then we have the

COLLISION EQUATION FOR COMPTON SCATTERING

Sum of energy-momentum vectors before collision

= Sum of energy-momentum vectors after collision

$$m_0(0,0,0,1)+h\nu(1,0,0,1) = \frac{m_0}{\sqrt{1-v^2}}(v\cos\beta, v\sin\beta, 0, 1) + h\nu'(\cos\alpha, \sin\alpha, 0, 1).$$

The first three components of this equation are called *conservation of (linear) momentum*, the last component is called *conservation of energy*, here $m_0 + h\nu = m_0/\sqrt{1-v^2} + h\nu'$.

The three nonzero momentum vectors form a triangle with two edges of lengths $h\nu, h\nu'$ and the angle $\alpha$ between them. The cosine theorem therefore gives the length (squared) of the third edge, which is the momentum (squared) of the particle. From the momentum squared we eliminate $v$ using the conservation of energy. This gives the relation between the deflection angle $\alpha$ and the frequency $\nu'$ (with $m_0, \nu$ given), or still simpler between $\alpha$ and the wave lengths $\lambda, \lambda'$ for **Compton Scattering**:

$$(h\nu)^2 + (h\nu')^2 - 2h^2\nu\nu'\cos\alpha = \frac{m_0^2}{1-v^2}v^2 = \frac{m_0^2}{1-v^2} - m_0^2 = (m_0 + h(\nu-\nu'))^2 - m_0^2$$

hence:

$$\frac{h}{m_0}(1-\cos\alpha) = (\nu-\nu')/(\nu\nu') = (\lambda'-\lambda).$$

The habit of mathematicians to take the speed of light as $c = 1$ and then drop $c$ from all formulas is not too popular among physicists. Certainly the most famous formula from all of physics, $E = mc^2$, that relates mass and energy, I have never seen without the speed of light being explicitly there. The connection between relativistic and classical energies is best illustrated by a power series expansion which also contains $c$ explicitly:

$$E = m(v)c^2 = \frac{m_0}{\sqrt{1-(v/c)^2}}c^2 = m_0c^2 + \frac{1}{2}mv^2 + \dots$$

We see: relativistic mass-energy equals restmass-energy plus classical kinetic energy plus higher order terms (in $(v/c)^2$).

An important notion in collision experiments is the *center of mass*. Consider a collection of particles (with restmas $\neq 0$, i.e. no photons) that fly with *different* constant velocities so that clocks tick differently on their world lines and mass depends on the velocity and therefore on the inertial system that we choose. How should one define their center of mass? Recall from classical mechanics that the sum of the momenta of a collection of particles equals the momentum of the center of mass (defined as sum of the masses times velocity of the center). This suggests for Special Relativity to add the momenta of the particles involved and write the result as (equivalent) mass of the center $M_c$ times a timelike unit vector $(v_c \in \mathbb{R}^3, 1)/\sqrt{1-v_c^2}$ that defines the inertial system of the center. Let $m_i$ be the rest masses of the particles involved and $v_i$ the velocities in some inertial system.

$$\sum_i \frac{m_i}{\sqrt{1-v_i^2}}(v_i, 1) =: \frac{M_c}{\sqrt{1-v_c^2}}(v_c, 1).$$

The timelike unit vector $(v_c, 1)/\sqrt{1-v_c^2}$ that defines the time axis of the center system is thus determined as an average of the timelike unit vectors of the world lines of the particles with their rest masses taken as weights. In particular, this definition is independent of the choice of the (inertial) coordinate system in which this equation is written. It describes the inertial system of the center, except that its origin, the center of mass *point*, is not yet defined. We switch to this center system and readjust notation: we still call the velocities of the mass points $v_i$. Using this center system means: $\sum_i m_i v_i/\sqrt{1-v_i^2} = 0$. Next we want to define the center point, more precisely, the world line of the center point. We intersect the world lines of the mass points $m_i$ with the simultaneity spaces of the center system (i.e., with the orthogonal complements of $v_c$). At center time $t = 0$ we call these points $(P_i, 0)$, at other times these intersection points are $(P_i(t), t) = (P_i, 0) + (v_i, 1) \cdot t$. Clearly, the average of the $P_i(t)$ with the relativistic weights $m_i/\sqrt{1-v_i^2}$ is independent of $t$. This average therefore defines the center of mass point in such a way that its world line is parallel to the time axis $\{(v_c, 1) \cdot t, t \in \mathbb{R}\}$ of the center system ($v_c = 0$ when in this system). Thus the minimal requirements for a reasonable definition are met. *Comparison with collision experiments shows that this center of mass is not changed in a collision.* Therefore the definition is not only mathematically but also physically reasonable.

We continue the physics introduction to Special Relativity by discussing the Maxwell equations after the following mathematical chapter.

## Pseudo-Riemannian Calculus

Covariant Derivative, Curvature Tensor, Einstein Tensor Jacobi Fields

The goal of covariant differentiation is to set up a machinery on manifolds that works as similar as possible to standard differentiation in a vector space. By definition of a manifold $M$ we need to write functions, vector fields etc on $M$ or maps between manifolds in terms of local coordinates. Of course one tries at first to differentiate the coordinate expressions in the same way as on a vector space. But problems arise: the second coordinate derivative of a function is NOT a bilinear form on $M$, and the derivative of a vector field in the direction of another vector field is NOT a vector field because the results depend on the coordinates in a way that does not occur in vector spaces: If one changes coordinates then the second derivative of the change of coordinates map interferes. If one has nothing more but a differentiable manifold one has to live with this inconvenience. However, if one has a Riemannian or a Pseudo-Riemannian metric on the manifold then one can do much better by adjusting differentiation to the given metric.

Notational Convention. Most books in analysis denote the first derivative of a map $F : \mathbb{R}^n \to \mathbb{R}^m$ in a way that later collides with other notations. Since the tangent spaces of manifolds $M$ are fairly universally denoted as $T_pM$ Serge Lang suggested to denote the first derivative of $F : M \to N$ by $TF : TM \to TN$, the derivative of $F$ at $p \in M$ by $TF|_p : T_pM \to T_{F(p)}N$ and the directional derivative of $F$ at $p$ in the direction of a tangent vector $X \in T_pM$ by $T_XF_{|_p} \in T_{F(p)}N$, or also $T_XF_{|_p} = TF_{|_p} \cdot X = TF_{|_p}(X)$. I did not run into any collisions of notation with this suggestion and I therefore adopted these conventions. Of course, if $F$ is given in terms of local coordinates $(x^1, \dots x^m)$ for $M$ and $(y^1, \dots y^n)$ for $N$ as $y^j = f^j(x^1, \dots, x^n)$ then these coordinates define bases for the tangent spaces of $M$ and $N$ and $TF|_p$ is with respect to these bases given by the Jacobi matrix $(\frac{\partial}{\partial x^i} f^j) = (TF)^j_i$.

Before I come to the covariant derivative I mention two earlier examples of successful definitions that avoided being disturbed by the second derivative of the change of coordinates. Consider two coordinate systems for the manifold $M$, call the change of coordinates map $\Psi$. I denote the coordinate expressions of two vector fields by $X, \tilde{X}, Y, \tilde{Y}$ with $\tilde{X} = T\Psi \cdot X$ and $\tilde{Y} = T\Psi \cdot Y$. We compute the derivative of $Y$ in the direction $X$ in the two coordinate systems: $T_XY$ and $T_{\tilde{X}}\tilde{Y} = T^2\Psi(X,Y) + T\Psi \cdot T_XY$. This shows how the second derivative $T^2\Psi$ interferes in an unwanted way. However, because of the symmetry $T^2\Psi(X,Y) = T^2\Psi(Y,X)$ this problem does not arise in the computation of the Lie bracket:

$$[\tilde{X}, \tilde{Y}] = T_{\tilde{X}}\tilde{Y} - T_{\tilde{Y}}\tilde{X} = T\Psi \cdot [X,Y] = T\Psi \cdot (T_XY - T_YX).$$

Similarly, a 1-form has the coordinate expressions $\omega, \tilde{\omega}$ with $\omega = \tilde{\omega} \cdot T\Psi$ and differentiation does not give a bilinear form because of $T_X\omega(.) = T_{\tilde{X}}\tilde{\omega} \cdot T\Psi(.) + \tilde{\omega} \cdot T^2\Psi(X,.)$. Again, the unwanted second derivative drops out when we compute the exterior derivative of $\omega$:

$$d\omega(X,Y) = (T_X\omega)(Y) - (T_Y\omega)(X) = d\tilde{\omega}(\tilde{X}, \tilde{Y}).$$

The Lie bracket and the exterior derivative are important notions of analysis, they were defined on manifolds long before the covariant derivative was invented. I hope the above computations show what kind of a problem has to be overcome.

The covariant derivative was developed in stages until the following characterization was reached:

> If a Riemannian or Pseudo-Riemannian metric $g(.,.)$ on a manifold$M$ is given then one has a uniquely defined **covariant derivative** $D_X Y$ (covariant means "compatible with coordinate changes", i.e., $T\Psi \cdot D_X Y = \tilde{D}_{\tilde{X}} \tilde{Y}$). $D_X Y$ is characterized by the following two properties (axioms):
>
> $$[X, Y] = D_X Y - D_Y X \qquad \text{(Symmetry)}$$
> $$T_X(g(Y, Z)) = g(D_X Y, Z) + g(Y, D_X Z) \qquad \text{(Product Rule)}$$

From the two axioms one obtains both, a coordinate expression for $D_X Y$ which does not show the compatibility with coordinate changes, and a coordinate independent expression, the *Koszul formula*. We do the invariant formula first:

$$\begin{aligned}
&T_Z(g(X, Y)) + T_Y(g(Z, X)) - T_X(g(Y, Z)) + g([X, Y], Z) - g([Z, X], Y) + g([Y, Z], X) = \\
&\quad g(D_Z X, Y) + g(X, D_Z Y) + g(D_Y Z, X) + g(Z, D_Y X) - g(D_X Y, Z) - g(Y, D_X Z) \\
&\quad + g([X, Y], Z) - g([Z, X], Y) + g([Y, Z], X) = \\
&\qquad\qquad 2g(D_Y Z, X).
\end{aligned}$$

The first line consists of coordinate independent terms, therefore the last line is coordinate independent. For the local expression we write out the left side of the product rule in coordinates: $T_X(g(Y, Z)) = (T_X g)(Y, Z) + g(T_X Y, Z) + g(Y, T_X Z)$ and do the same $+, +, -$ cyclic sum as for the Koszul formula. We simplify using $[X, Y] = T_X Y - T_Y X$ etc. and obtain the local expression for $2g(D_Y Z, X)$:

$$2g(D_Y Z, X) = 2g(T_Y Z, X) + (T_Z g)(X, Y) + (T_Y g)(Z, X) - (T_X g)(Y, Z)$$

or, with the definition of the Christoffel symbols (note the symmetry):

$$g(\Gamma(Y, Z), X) := (T_Z g)(X, Y) + (T_Y g)(Z, X) - (T_X g)(Y, Z) = g(\Gamma(Z, Y), X),$$
$$D_Y Z = T_Y Z + \Gamma(Y, Z).$$

If one wants to see indices one has to use bases, either some orthonormal moving frame $\{e_1, \ldots, e_n\}$ or the basis coming from the coordinates $e_j := \frac{\partial}{\partial x^j}$. Then

$$Y = \sum_j y^j e_j, \quad \omega(e_j) =: \omega_j, \quad \omega(Y) = \sum_j y^j \omega_j, \quad g(e_i, e_j) =: g_{ij}, \quad \Gamma(e_j, e_k) = \sum_i \Gamma^i_{jk} e_i.$$

The so called *Einstein sum convention* omits all $\sum$-signs and assumes that pairs of lower and upper indices are summation indices.

EXAMPLE. If the metric $g$ on a submanifold is induced from the metric of the surrounding space then the covariant derivative of the submanifold metric is the *tangential component* of the covariant derivative in the surrounding space — because this tangential component satisfies the two axioms above.

EXAMPLE. We call a vector field along a curve $c$ *parallel* if its covariant derivative vanishes: $D_{\dot{c}}Y = 0$. In local coordinates this is a first order linear differential equation $\dot{Y}(t) + \Gamma_{|_{c(t)}}(\dot{c}(t), Y(t)) = 0$. Every initial vector $Y(0) \in T_{c(0)}M$ extends to a parallel field, and a basis of orthonormal initial vectors extends to an orthonormal basis of parallel fields along the curve $c$.

Next we have to extend the definitions to other objects, we want to differentiate forms, endomorphisms, in general: tensors. These objects have in common that we can represent them by a bunch of components as soon as we have bases in the tangent spaces involved. In the Euclidean situation we call a form or an endomorphism field or any tensor field

**parallel** if its components with respect to a parallel basis are **constant.**

Of course, arbitrary tensor fields along a curve are linear combinations of parallel fields with functions as coefficients. These are differentiated by differentiating the coefficient functions as in the standard vector space situation. Thus we have defined directional derivatives of tensor fields. Finally, as in the standard situation, if these directional derivatives are continuous then the result of the differentiation depends *linearly* on the direction vector. This linear map is then called the covariant differential of the tensor field.

Since this whole sequence of definitions is completely the same as in the standard situation (i.e., in a vector space instead of in a manifold) we have of course those same differentiation rules to which we are used: linearity, product rule, chain rule, computations in terms of partial derivatives. So, why do differential geometry computations look so different from standard analysis computations? The reason is that, of course, we always choose parallel bases in standard computations and we will see that these are in general not available on a manifold (except along curves). Now, if the components that we differentiate are NOT with respect to a parallel basis, then we do not get the derivative of the tensor before we correct for the nonvanishing derivative of the basis fields: If $\omega_j(t) := \omega_{|_t}(e_j(t))$ then $\dot{\omega}_j(t) := \dot{\omega}_{|_t}(e_j(t)) + \omega_{|_t}(\dot{e}_j(t))$. In many textbooks this is expressed by saying: the derivative of an endomorphism field $A$ is *defined* as $(D_X A)(Y) := D_X(A(Y)) - A(D_X Y)$. If this were indeed the definition then we would have a huge difference from our standard theory and we could not really expect differentiation rules to be similar. But as I explained, we know the derivative of tensor fields before this formula and the formula is the computational way to deal with non-parallel bases.

Here is another point where standard and covariant differentiation are closer than it often looks. Let the above endomorphism field be the covariant differential of a vector field: $A \cdot Y := D_Y Z$. In this case we have to distinguish two different second derivatives, the *iterated* second derivative $D_X(D_Y Z)$ (which appears more frequently in printed computations) and the tensorial second derivative $D^2_{X,Y}Z = D_X(D_Y Z) - D_{D_X Y}Z$ which is closer to the standard second derivative and which is tensorial in $X$ and $Y$, i.e. we have linearity with functions $f, g$ as coefficients: $D^2_{fX,gY}Z = fgD^2_{X,Y}Z$.

We have gone through almost all properties of standard differentiation and made the covariant derivative look the same. There is only one property left namely the symmetry of second and higher derivatives, and these symmetries are not shared by the covariant derivative. We compute the local expression of $D^2_{X,Y}Z - D^2_{Y,X}Z$ and find:

$$D^2_{X,Y}Z - D^2_{Y,X}Z = (T_X\Gamma)(Y,Z) - (T_Y\Gamma)(X,Z) + \Gamma(X,\Gamma(Y,Z)) - \Gamma(Y,\Gamma(X,Z)).$$

This is a very remarkable result: The left side is an invariant expression (it is independent of the coordinate system), the right side quite unexpectedly does not depend on the derivatives of $Z$, it is tensorial also in the argument $Z$! Of course such a surprise gives rise to a definition:

The **Riemann Curvature Tensor:**

$$R(X,Y)Z := D^2_{X,Y}Z - D^2_{Y,X}Z.$$

Note however that the covariant hessian of a function does not feel the curvature, it is still symmetric. Observe that for the first derivative of a function there is no difference between standard derivative and covariant derivative, $T_X f = D_X f$.

| | |
|---|---|
| Standard Hessian: | $T_X(T_Y f) - T_{T_X Y} f =: \text{hess}_{std} f(X,Y)$ |
| Covariant Hessian: | $T_X(T_Y f) - T_{D_X Y} f =: \text{hess}_{cov} f(X,Y)$ |
| | $\text{hess}_{cov} f(X,Y) - \text{hess}_{std} f(X,Y) = -T_{\Gamma(X,Y)} f,$ |
| Hessian Symmetry: | $\text{hess}_{cov} f(X,Y) = \text{hess}_{cov} f(Y,X).$ |

When the skew symmetric part of the second derivative is applied to a product of tensor fields $A, B$ then the first derivatives drop out. Let $A, B$ be tensor fields for which a product $A \cdot B$ is defined. Then we have a

**Product Rule for** $(D^2_{X,Y} - D^2_{Y,X})$**:**

$$(D^2_{X,Y} - D^2_{Y,X})(A \cdot B) = ((D^2_{X,Y} - D^2_{Y,X})A) \cdot B + A \cdot (D^2_{X,Y} - D^2_{Y,X})B.$$

Example for a form $\omega$ and a vector field $Z$:

$$0 = (D^2_{X,Y} - D^2_{Y,X})(\omega \cdot Z) = ((D^2_{X,Y} - D^2_{Y,X})\omega) \cdot Z + \omega \cdot R(X,Y)Z$$

Example for an endomorphism field $A$ and a vector field $Z$:

$$R(X,Y)(A \cdot Z) = (D^2_{X,Y} - D^2_{Y,X})(A \cdot Z) = ((D^2_{X,Y} - D^2_{Y,X})A) \cdot Z + A \cdot R(X,Y)Z.$$

Example for the metric $g(.,.)$ and two vector fields $V, W$:

$$0 = (D^2_{X,Y} - D^2_{Y,X})g(V,W) = g((D^2_{X,Y} - D^2_{Y,X})V, W) + g(V, R(X,Y)W).$$

For working with the curvature tensor it is important to understand its symmetries. In the following list the first line is true by definition, the second follows from the local formula, the third one rephrases the last example to the product rule for $(D^2_{X,Y} - D^2_{Y,X})$ and the fourth line follows from the first three:

**Symmetries of the curvature tensor:**

| | |
|---|---|
| Skew Symmetry in the first pair | $R(X,Y)Z = -R(Y,X)Z$ |
| 1. Bianchi Identity | $R(X,Y)Z + R(Y,Z)X + R(Z,X)Y = 0$ |
| Skew Symmetry in the second pair | $g(R(X,Y)V,W) = -g(R(X,Y)W,V)$ |
| Symmetry in both pairs | $g(R(X,Y)V,W) = g(R(V,W)X,Y)$ |

**Hypersurface Theory** is the same as in the Riemannian case, except for signs related to the normal $N$. For some manifold $M$ let $F : M \to \mathbb{R}^n, \langle\langle .,. \rangle\rangle$ be a hypersurface immersion such that $\langle\langle N,N\rangle\rangle = \pm 1$ and does not change sign. Since we assume that the metric is induced from $\mathbb{R}^n$ we differentiate $\langle\langle TF(Y), TF(Z)\rangle\rangle - g(Y,Z) = 0$ and since $F = F^1, \ldots, F^n$ is a collection of $n$ functions the above definitions apply: $D^2F(X,Y) = T_X(T_Y F) - T_{D_X Y}F$. Hence

$$\langle\langle D^2F(X,Y), T_Z F\rangle\rangle + \langle\langle T_Y F, D^2F(X,Z)\rangle\rangle = 0.$$

Next we do the same $+,+,-$ cyclic computation as for the Koszul formula and, noting the symmetry $D^2F(X,Y) = D^2F(Y,X)$ we get a first result: $D^2F(Y,Z)$ is normal:

$$2\langle\langle T_X F, D^2F(Y,Z)\rangle\rangle = 0, \text{ hence } D^2F(Y,Z) = \langle\langle D^2F(Y,Z), N\rangle\rangle / \langle\langle N,N\rangle\rangle \cdot N.$$

Recall the definition of the **shape operator** (or Weingarten map, or second fundamental tensor) and differentiate $0 = \langle\langle N, T_Y F\rangle\rangle$ to relate the shape operator and $D^2F(X,Y)$:

$$T_Y N =: TF \cdot S \cdot Y$$
$$0 = \langle\langle T_X N, T_Y F\rangle\rangle + \langle\langle N, D^2F(X,Y)\rangle\rangle, \text{ or } g(SX,Y) = -\langle\langle N, D^2F(X,Y)\rangle\rangle.$$

In particular, the shape operator is $g$-symmetric. Next, differentiate the definition of $S$, note the normal and the tangential component of the result and get the Codazzi equation:

$$D^2_{X,Y}N = D^2_{Y,X}N = D^2F(X,SY) + TF((D_X S)Y)$$

**Codazzi Equation:** $\quad (D_X S)Y = (D_Y S)X.$

Finally differentiate $g(SY,Z)/\langle\langle N,N\rangle\rangle \cdot N = -D^2F(Y,Z)$ observe tangential and normal components and get the Gauss equation by observing the product rule:

$$0 = (D^2_{X,Y} - D^2_{Y,X})(T_Z F) = (D^3_{X,Y,Z} - D^3_{Y,X,Z}F) + TF \cdot R(X,Y)Z$$

**Gauss Equation:** $\quad g(SY,Z)SX - g(SX,Z)SY = \langle\langle N,N\rangle\rangle R(X,Y)Z.$

From the full curvature tensor one defines, by taking a trace, a simpler tensor that will be important for formulating the Einstein equations.

**Definition of the Ricci tensor:** $g(Ric\,Y,Z) = ric(Y,Z) := \sum_i g(R(Y,e_i)e_i,Z)/g(e_i,e_i)$, where $\{e_1,\ldots,e_n$ is an orthogonal basis (not necessarily orthonormal). The Ricci tensor is $g$-symmetric because of the symmetries of the curvature tensor - but only for positive definite $g$ does this imply the existence of a basis consisting of eigenvectors of $Ric$.

The Codazzi equation leads to somewhat analogous equations for the curvature tensor and for the Ricci tensor from which one gets the following

**Constancy Theorems**

| | |
|---|---|
| Umbilicity theorem | $S = f(p)id \Rightarrow f = const.$ |
| Schur's theorem, dim $> 2$ | $R(X,Y)Z = f(p)(g(Y,Z)X - g(X,Z)Y) \Rightarrow f = const.$ |
| Einstein metric, dim$> 2$ | $ric = f(p)g \Rightarrow f = const.$ |

Before the proofs I derive the identities for the other tensors. Differentiate the Gauss equation and observe, that the cyclic sum over $U,X,Y$ gives zero (because of Codazzi):

$$g((D_US)Y,Z)SX - g((D_US)X,Z)SY + g(SY,Z)(D_US)X - g(SX,Z)(D_US)Y$$
$$= \langle\langle N,N\rangle\rangle(D_UR)(X,Y)Z.$$

**2. Bianchi Identity:** $\quad (D_UR)(X,Y)Z + (D_XR)(Y,U)Z + (D_YR)(U,X)Z = 0.$

This short proof applies only to curvature tensors of hypersurfaces. The general case can be obtained by suitably applying differentiation rules. Consider, for a given vector field $Z$ the definition of the curvature tensor, $(D^2_{X,Y} - D^2_{Y,X})Z = R(X,Y)Z$, as an equation between vector valued twoforms and differentiate once more. (Another way to justify the following is, to assume that the fields $X$ and $Y$ are parallel in direction $U$). We obtain:

$$(D^3_{U,X,Y} - D^3_{U,Y,X})Z = (D_UR)(X,Y)Z + R(X,Y)D_UZ.$$

To obtain another commutation formula apply the product rule example for endomorphism fields to the endomorphism $A\cdot X = D_XZ$. Note $((D^2_{U,Y} - D^2_{Y,U})A)\cdot X = (D^3_{U,Y,X} - D^3_{Y,U,X})Z$. Then the quoted product rule gives

$$(D^3_{U,Y,X} - D^3_{Y,U,X})Z = R(U,Y)D_XZ - D_{R(U,Y)X}Z.$$

These two commutation formulas combine to

$$(D^3_{U,X,Y} - D^3_{Y,U,X})Z = (D_UR)(X,Y)Z - R(Y,U)D_XZ + R(X,Y)D_UZ - D_{R(U,Y)X}Z.$$

Cyclic permutation over $(U,X,Y)$ and summation kills most terms, only the second Bianchi identity remains. Q.E.D.

To see what the second Bianchi identity implies for the Ricci tensor, we compute the divergence of the Ricci tensor and the derivative of its trace. From the definition we have $g((D_URic)\cdot Y,Z) = \sum_i g((D_UR)(Y,e_i)e_i,Z)/g(e_i,e_i)$, hence

$$g(\operatorname{div}(Ric),Z) = \sum_{i,j} g((D_{e_j}R)(e_j,e_i)e_i,Z)/(g(e_i,e_i)g(e_j,e_j)).$$

Again from the definition we compute trace Ricci:

$$T_Z \text{trace}(Ric) = \sum_{i,j} g((D_Z R)(e_j, e_i)e_i, e_j)/(g(e_i, e_i)g(e_j, e_j)).$$

The second Bianchi identity and the curvature symmetries imply

$$2g(\text{div}(Ric), Z) = T_Z \text{trace}(Ric).$$

Therefore we can define the

**Divergence free Einstein tensor:** $\quad G := Ric - \frac{1}{2}\text{trace}(Ric) \cdot id.$

The proofs of the constancy results are now immediate, e.g. $S = f(p)id \Rightarrow (D_X S)Y = (T_X f)Y$ and if we use the Codazzi equation with two independent vectors $X, Y$ we get $Tf = 0$.
Similarly for Schur's theorem: $R(X,Y)Z = f(p)(g(Y,Z)X - g(X,Z)Y) \Rightarrow$
$$(D_U R)(X,Y)Z = (T_U f)(g(Y,Z)X - g(X,Z)Y).$$
If one chooses $X, Y, Z \perp U,\ Z \perp X,\ Y = Z$ then only one term remains in the 2. Bianchi identity, $0 = (T_U f)g(Y,Y)X$, and $f$ is constant.
Finally the Ricci case: $\quad Ric = f(p) \cdot id \Rightarrow$
$$D_Z Ric = T_Z f \cdot id,\ T_Z(\text{trace}Ric) = nT_Z f,\ 2g(\text{div}(Ric), Z) = 2T_Z f.$$
If $n > 2$ then $\text{div}(G) = 0$ implies $T_Z f = 0$, so $f$ is constant.

## An important tool: The Jacobi Equation

Recall that, any time one has a 1-parameter family of solutions of some nonlinear (differential) equation, then one can differentiate the family with respect to its parameter to obtain an object *that is a solution of a linear (differential) equation*. A family of geodesics is a family of solutions of the geodesic equation $\frac{D}{dt}\dot{c} = 0$ and differentiation with respect to the family parameter gives a vector field along each geodesic. Because of the general statement above this vector field must solve a linear second order ODE. We should expect that the coefficients of the equation are some geometric invariant. This ODE is as important for the geometry as the derivative is for the study of a function. This ODE is called the *Jacobi equation*. For its derivation we have to commute differentiations in different directions, therefore the curvature tensor must show up.

Let $c(s,t)$ be a family of geodesics, $s$ is the family parameter and $t \mapsto c(s,t)$ are geodesics, i.e. $\frac{D}{dt}\frac{d}{dt}c(s,t) = 0$. We abbreviate $\dot{c}(s,t) := \frac{d}{dt}c(s,t)$ and $c'(s,t) := \frac{d}{ds}c(s,t)$. First observe the symmetry

$$\frac{D}{dt}\frac{d}{ds}c(s,t) = \frac{d}{dt}\frac{d}{ds}c(s,t) + \Gamma(\dot{c}, c') = \frac{d}{ds}\frac{d}{dt}c(s,t) + \Gamma(\dot{c}, c') = \frac{D}{ds}\frac{d}{dt}c(s,t)$$

which implies for every vector field $v(s,t)$ along $c(s,t)$

$$\frac{D}{dt}\frac{D}{ds}v(s,t) - \frac{D}{ds}\frac{D}{dt}v(s,t) = D^2_{\dot c, c'}v(s,t) - D^2_{c',\dot c}v(s,t) = R(\dot c, c')v(s,t).$$

We apply this to $\frac{D}{ds}\frac{D}{dt}\frac{d}{dt}c(s,t) = 0, \text{i.e., } v(s,t) = \dot c(s,t),$ and obtain

$$0 = \frac{D}{dt}\frac{D}{ds}\frac{d}{dt}c(s,t) + R(c',\dot c)\dot c = \frac{D}{dt}\frac{D}{dt}c'(s,t) + R(c',\dot c)\dot c$$

which is the looked for linear second order ODE for $c'$, called the **Jacobi Equation**

Note that $J \mapsto R(J,\dot c)\dot c$ is a $g$-symmetric operator. It has $\dot c$ as eigenvector with eigenvalue 0 and it maps the orthogonal complement $\{\dot c\}^\perp$ into itself. In case $g$ has the signature of Special Relativity and $t \mapsto c$ is a timelike geodesic then $g$ is positive definite on the orthogonal complement $\{\dot c\}^\perp$. Therefore one can estimate $J \mapsto R(J,\dot c)\dot c$ by the smallest eigenvalue $\delta$ from below and by the largest eigenvalue $\Delta$ from above, on $\{\dot c\}^\perp$:

$$\delta \cdot g(J,J) \le g(R(J,\dot c)\dot c, J) \le \Delta \cdot g(J,J).$$

Because the curvature tensor has so many indices some people believe that the Jacobi equation is more complicated than the geodesic equation. To weaken this belief somewhat I prove an important inequality for Jacobi fields $J \perp \dot c$ and satisfying $J(0) = 0$:

$$\begin{aligned}
\frac{d}{dt}|J| &= g(J, \frac{D}{dt}J)/|J|, \\
\frac{d}{dt}\frac{d}{dt}|J| &= g(J, \frac{D}{dt}\frac{D}{dt}J)/|J| + g(\frac{D}{dt}J, \frac{D}{dt}J)/|J| - g(J, \frac{D}{dt}J)^2/|J|^3 \\
&\ge -g(J, R(J,\dot c)\dot c)/|J| \qquad \text{by Schwarz inequality} \\
&\ge -\Delta \cdot |J|.
\end{aligned}$$

This inequality is used to show that the function $f(t) := |J(t)|/\mathbf{s}_\Delta(t)$ is increasing (the definition of $\mathbf{s}_\Delta()$ is $\frac{d}{dt}\frac{d}{dt}\mathbf{s}_\Delta(t) + \Delta \mathbf{s}_\Delta(t) = 0,\ \mathbf{s}_\Delta(0) = 0,\ \frac{d}{dt}\mathbf{s}_\Delta(0) = 1$):

$$\begin{aligned}
\dot f(t) &= \big(\frac{d}{dt}|J| \cdot \mathbf{s}_\Delta(t) - |J| \cdot \dot{\mathbf{s}}_\Delta(t)\big)/\mathbf{s}_\Delta(t)^2 \\
&= \big(\int_0^t (\frac{d}{dt}\frac{d}{dt}|J| \cdot \mathbf{s}_\Delta(t) - |J| \cdot \ddot{\mathbf{s}}_\Delta(t))dt\big)/\mathbf{s}_\Delta(t)^2 \\
&\ge 0
\end{aligned}$$

Since by l'Hospital $f(0) = |J'(0)|$ we have one of the **Rauch estimates:**

$$|J'(0)| \cdot \mathbf{s}_\Delta(t) \le |J(t)|.$$

If $\Delta \le 0$ this says that $J$ is growing at least linearly and if $\Delta > 0$ this implies that $J$ has no zero in $(0, \pi/\sqrt{\Delta})$, so that there are no conjugate points in this interval.

## Special Relativity II

Maxwell's Equations, Hodge-$*$, Conformal Invariance, Plane Waves, Lorentz Force, Aberration of Light.

Maxwell's equations are included in this introduction to Special Relativity for the following reasons: Practically all astronomical information reaches us via electromagnetic waves; the most important cosmological models are conformally flat. therefore we will use the conformal invariance of the Maxwell equations and how field strengths change under conformal changes of the metric; the measured electromagnetic fields depend on the considered solution to the equations and on the observer, but only on the *rest frame* of the observer – such observers I will call *infinitesimal observers* and they will help us to connect theory and experiment also in other situations.

We will write the indefinite scalar product as

$$\langle\langle X, X\rangle\rangle = (x^1)^2 + (x^2)^2 + (x^3)^2 - (x^4)^2,\ x^4 = c \cdot t \qquad \text{(the } \textit{Lorentz form}\text{)}$$

The dual basis is $\{\vec{e}_1, \ldots, \vec{e}_4\}$ with $dx^i(\vec{e}_j) = \delta^i_j$.

Threedimensional formulation of Maxwell's Equations:

$$\vec{D} = \epsilon_0 \vec{E}\,, \quad \vec{B} = \mu_0 \vec{H}\,, \quad \epsilon_0\mu_0 = c^{-2}$$

homogenous equations $\quad \operatorname{rot} \vec{E} = -\frac{d}{dt}\vec{B}\,, \quad \operatorname{div} \vec{B} = 0$

non-homogenous equations $\quad \operatorname{rot} \vec{H} = \vec{j} + \frac{d}{dt}\vec{D}\,, \quad \operatorname{div} \vec{D} = \rho.$

Next we define a twoform, the *Faradayform* $F$, from $\vec{E}, \vec{B}$ and show that the homogenous Maxwell equations can be expressed as $dF = 0$ (which is a very coordinate independent formulation). We introduce the components of $\vec{E}, \vec{B}$ with respect to the above basis, $\vec{E} = \sum_i E_i\vec{e}_i,\ \vec{B} = \sum_i B_i\vec{e}_i$ and define the

**Faraday Form**

$$F := (E_1 dx^1 + E_2 dx^2 + E_3 dx^3) \wedge dx^4 + B_1 dx^2 \wedge dx^3 + B_2 dx^3 \wedge dx^1 + B_3 dx^1 \wedge dx^2$$

The matrix associated to $F$ is

$$F(\vec{e}_i, \vec{e}_j) = \begin{pmatrix} 0 & B_3 & -B_2 & E_1 \\ -B_3 & 0 & B_1 & E_2 \\ B_2 & -B_1 & 0 & E_3 \\ -E_1 & -E_2 & -E_3 & 0 \end{pmatrix}$$

Vice versa, if such a twoform is given then we define ,
$E_i := F(\vec{e}_i, \vec{e}_4), i = 1, 2, 3, \quad B_k := F(\vec{e}_i, \vec{e}_j),\ (i, j, k)$ a cyclic permutation of $(1, 2, 3)$, and the fields $\vec{E} = \sum_i^3 E_i\vec{e}_i,\ \vec{B} = \sum_i^3 B_i\vec{e}_i$ depend only on the tangent vector $\vec{e}_4$ to the world line of the observer, since a Lorentz transformation that preserves $\vec{e}_4$ is a usual orthogonal transformation of the rest space $(\vec{e}_4)^\perp$ of the observer, so $\vec{E}, \vec{B}$ do not change.

We compute $dF$

$$dF = \sum_i dE_i \wedge dx^i \wedge dx_4 \quad + \sum_{(i,j,k)=(1,2,3)} dB_i \wedge dx^j \wedge dx^k,$$

where the subcript under the second sum means that we sum over all *cyclic* permutations of $(1,2,3)$. Of course, for any function, $df = \sum \frac{\partial}{\partial x^i} f\, dx^i$. The terms in the second sum that do not contain $dx^4$ are all multiples of $dx^i \wedge dx^j \wedge dx^k = dx^1 \wedge dx^2 \wedge dx^3$ and the coefficients add up to $\sum \frac{\partial}{\partial x^i} B_i = \operatorname{div} \vec{B}$. All other terms contain $dx^4$. The first sum is

$$\sum_i \left(\frac{\partial}{\partial x^j} E_i dx^j + \frac{\partial}{\partial x^k} E_i dx^k\right) \wedge dx^i \wedge dx^4 = \sum_i \left(-\frac{\partial}{\partial x^j} E_i dx^i \wedge dx^j + \frac{\partial}{\partial x^k} E_i dx^k \wedge dx^i\right) \wedge dx^4$$

We assume again that $(i,j,k)$ is a cyclic permutation of $(1,2,3)$ and we recall that $(\operatorname{rot} \vec{E})_j = \frac{\partial}{\partial x^k} E_i - \frac{\partial}{\partial x^i} E_k$. Therefore one can reorganize the first sum also as a sum over cyclic permuatations;

$$\sum_{(i,j,k)=(1,2,3)} (\operatorname{rot} \vec{E})_i dx^j \wedge dx^k \wedge dx^4.$$

Therefore we have the homogenous Maxwell equations expressed as $dF = 0$:

$$dF = 0 \qquad \Leftrightarrow \qquad \operatorname{div} \vec{B} = 0, \;\; \operatorname{rot} \vec{E} + \frac{\partial}{\partial x^4} \vec{B} = 0.$$

The first half of the Maxwell equations therefore means: $F$ is a closed twoform, and this has nothing to do with the Lorentz scalar product that we want to use. The second set of equations does depend on the metric, but I postpone discussing them and first explain how physical situations are described with the help of the Faraday form.

1. Example: Charged Wire. In the rest system of the wire (along the x-axis) we assume a charge density of $\rho = 1$ per unit length. There we observe no magnetic field and the electric field is orthogonal to the wire (and of strength $1/r$). Therefore we have:

$$\vec{E} = \left(0, \frac{y}{y^2+z^2}, \frac{z}{y^2+z^2}\right), \qquad \vec{B} = 0$$

$$F = \left(\frac{y dy}{y^2+z^2} + \frac{z dz}{y^2+z^2}\right) \wedge dt = \frac{1}{2}(\mathrm{d}\log r^2) \wedge dt$$

Clearly $dF = 0$ and we check the other equations later. Now consider a second observer that flies in the x-direction with velocity $v$, what does he see? First, his unit timelike vector is $\vec{f}_4 = (v\vec{e}_1 + \vec{e}_4)/\sqrt{1-v^2}$ and a convenient basis in his restspace is
$\vec{f}_1 := (\vec{e}_1 + v\vec{e}_4)/\sqrt{1-v^2}, \vec{f}_2 := \vec{e}_2, \vec{f}_3 := \vec{e}_3$.
We plug this frame into the Faraday form and find

$$(\vec{E})_{new} : F(\vec{f}_1, \vec{f}_4) = 0, \; F(\vec{f}_2, \vec{f}_4) = \frac{1}{\sqrt{1-v^2}} \frac{y}{y^2+z^2}, \; F(\vec{f}_3, \vec{f}_4) = \frac{1}{\sqrt{1-v^2}} \frac{z}{y^2+z^2},$$

$$(\vec{B})_{new} : F(\vec{f}_2, \vec{f}_3) = 0, \; F(\vec{f}_3, \vec{f}_1) = \frac{v}{\sqrt{1-v^2}} \frac{z}{y^2+z^2}, \; F(\vec{f}_1, \vec{f}_2) = \frac{v}{\sqrt{1-v^2}} \frac{-y}{y^2+z^2},$$

First, the second observer sees a stronger electric field. Most important, this does agree with observation. But, from our geometric point of view, we would actually expect this before the experiment, why? Take the world lines of the electrons on one unit of length on the wire. Their world lines are parallel to the time axis of the first observer and they carry the charge $\rho$. These worldlines intersect the restspace of the second observer in a segment that is shorter by the factor $\sqrt{1-v^2}$, therefore the charge density is larger by this factor and the electric field is accordingly stronger. – Secondly, the magnetic field is, in agreement with observations, proportional to the {*charge per length times the velocity*}, i.e. proportional to the electric current in the wire. The magnetic field lines are circles around the wire. The size of the field decreases as $1/r$.

One should pause to contemplate this result for a moment: One writes down the Faraday form in the inertial system in which the form is simplest. Then one obtains the electric and magnetic fields of any observer by plugging its 4-dimensional rest frame into the form. The procedure could hardly be simpler.

2. Example: Point Charge. In the rest system of a point charge of size $q$ we have no magnetic field and the radially symmetric electric Coulomb field:

$$\vec{E} = \frac{q}{r^3}\cdot(x,y,z), \quad \vec{B}=0 \qquad\qquad \text{hence}$$
$$F = \frac{q}{r^3}(xdx+ydy+zdz)\wedge dt = \frac{q}{r^2}dr\wedge dt.$$

Consider an observer that rotates around the charge with velocity $v$ on a circle of radius $r$ at height $z=h$. This non-inertial observer has the

world line $\qquad c(t) = (r\cos(v/r\cdot t), r\sin(v/r\cdot t), h, t)$

and the rest frame at time $t=0$ is:

$$\vec{f}_1 = \vec{e}_1,\ \vec{f}_2 = (\vec{e}_2 + v\vec{e}_4)/\sqrt{1-v^2},\ \vec{f}_3 = \vec{e}_3,\ \vec{f}_4 = \dot{\vec{c}}(0) = (\vec{v}e_2 + \vec{e}_4)/\sqrt{1-v^2}.$$

Although this observer is not inertial we obtain the electromagnetic field that he experiences by plugging its rest frame at time $t$ into the Faraday form. At $t=0$ we obtain:

$$E_1 = F(\vec{f}_1,\vec{f}_4) = \frac{x}{r^3}\frac{q}{\sqrt{1-v^2}}, \quad E_2 = 0, \quad E_3 = F(\vec{f}_3,\vec{f}_4) = \frac{z}{r^3}\frac{q}{\sqrt{1-v^2}},$$
$$B_1 = F(\vec{f}_2,\vec{f}_3) = 0, \quad B_2 = F(\vec{f}_3,\vec{f}_1) = 0, \quad B_3 = F(\vec{f}_1,\vec{f}_2) = \frac{x}{r^3}\frac{qv}{\sqrt{1-v^2}}.$$

It is more interesting to let the point charge (or the charged wire) rotate around the observer, but we cannot exchange the observer and the charge as in the first example since not both of them are inertial. However, in many interesting situations the charges move slowly (e.g. for a current in a wire the speed is millimeter per second). Therefore we can use the above computation to obtain the field created by a rotating charge (a current in a circular wire) for an observer at rest (on the rotation axis or, with more work, off the axis). I will not pursue this further because the physicists solve the Maxwell equations from scratch for those other situations. My goal has been to explain how *different* observers see the electromagnetic fields coming from the *same* Faraday form.

We return now to the non-homogenous Maxwell equations, in particular, how are they expressed in terms of the Faraday form? For that we need to discuss the

$$\textbf{Hodge-map} \quad * : \Lambda^k \to \Lambda^{4-k}.$$

First we define the induced scalar products on $\Lambda^1, \ldots, \Lambda^4$. On $\Lambda^1$ we use the dual basis $dx_1, \ldots, dx_4,\ dx^i(\vec{e}_j) = \delta^i_j$ and we define

$$\omega, \mu \in \Lambda^1 \Rightarrow \quad \langle\langle \omega, \mu \rangle\rangle := \sum_{i=1}^{4} \frac{\omega(\vec{e}_i)\mu(\vec{e}_i)}{\langle\langle \vec{e}_i, \vec{e}_i \rangle\rangle} \quad \text{hence} \quad \langle\langle dx^i, dx^j \rangle\rangle = \langle\langle \vec{e}_i, \vec{e}_j \rangle\rangle,$$

in particular, the scalar product on $\Lambda^1$ has the same signature as the given Lorentz form. Similarly for the other cases:

$$\omega, \mu \in \Lambda^2 \Rightarrow$$
$$\langle\langle \omega, \mu \rangle\rangle := \sum_{i<k} \frac{\omega(\vec{e}_i, \vec{e}_k)\mu(\vec{e}_i, \vec{e}_k)}{\langle\langle \vec{e}_i, \vec{e}_i \rangle\rangle \langle\langle \vec{e}_k, \vec{e}_k \rangle\rangle} \quad \text{hence} \quad \langle\langle dx^i \wedge dx^k, dx^i \wedge dx^k \rangle\rangle = \langle\langle \vec{e}_i, \vec{e}_i \rangle\rangle \langle\langle \vec{e}_k, \vec{e}_k \rangle\rangle,$$
$$\omega, \mu \in \Lambda^3 \Rightarrow$$
$$\langle\langle \omega, \mu \rangle\rangle := \sum_{i<k} \frac{\omega(\vec{e}_i, \vec{e}_j, \vec{e}_k)\mu(\vec{e}_i, \vec{e}_j, \vec{e}_k)}{\langle\langle \vec{e}_i, \vec{e}_i \rangle\rangle \langle\langle \vec{e}_j, \vec{e}_j \rangle\rangle \langle\langle \vec{e}_k, \vec{e}_k \rangle\rangle} \quad \text{hence}$$
$$\langle\langle dx^i \wedge dx^j \wedge dx^k,\ dx^i \wedge dx^j \wedge dx^k \rangle\rangle = \langle\langle \vec{e}_i, \vec{e}_i \rangle\rangle \langle\langle \vec{e}_j, \vec{e}_j \rangle\rangle \langle\langle \vec{e}_k, \vec{e}_k \rangle\rangle,$$
$$\Lambda^4 \text{ has the canonical basis } vol := dx_1 \wedge dx_2 \wedge dx_3 \wedge dx_4, \quad \langle\langle vol, vol \rangle\rangle = -1.$$

These indefinite scalar products are independent of the used orthonormal basis and they are non-degenerate in all four cases. Therefore they give isomorphisms between the $\Lambda^k$ and their dual spaces $(\Lambda^k)^*$ via $\omega \mapsto \langle\langle \omega, .\ \rangle\rangle$. Also, the $\wedge$-product gives an isomorphism between $\Lambda^{4-k}$ and $(\Lambda^k)^*$: For $\mu \in \Lambda^{4-k}$ define $l_\mu \in (\Lambda^k)^*$ by $\omega \wedge \mu = l_\mu(\omega) \cdot vol$. Now we can define the Hodge-$*$-map independent of the used orthonormal bases as follows:

$$* : \Lambda^k \to \Lambda^{4-k} \quad \text{by representing for each } \mu \in \Lambda^k \text{ the map}$$
$$\omega \in \Lambda^k \to \langle\langle \omega, \mu \rangle\rangle \cdot vol \quad \text{as} \quad \omega \wedge *\mu := \langle\langle \omega, \mu \rangle\rangle \cdot vol.$$

Of course, the Hodge-$*$ depends on the Lorentz form. But one can check, that it behaves simple if one changes the Lorentz form conformally, i.e., $\langle\langle ., . \rangle\rangle$ changed to $f^2(p) \cdot \langle\langle ., . \rangle\rangle$. Namely, on $\Lambda^2$ the Hodge-$*$ is invariant under conformal changes and on $\Lambda^1, \Lambda^3, \Lambda^4$ it is just a conformal map.

Because of the many minus-signs in the formulas it is convenient to have a list of how Hodge-$*$acts on the usual bases:

$$\begin{array}{ccc} * : \Lambda^1 & \to & \Lambda^3 \\ dx^1 & \mapsto & dx^2 \wedge dx^3 \wedge dx^4 \\ dx^2 & \mapsto & dx^3 \wedge dx^1 \wedge dx^4 \\ dx^3 & \mapsto & dx^1 \wedge dx^2 \wedge dx^4 \\ dx^4 & \mapsto & dx^1 \wedge dx^2 \wedge dx^3 \\ & & \\ * : \Lambda^3 & \to & \Lambda^1 \\ *(*(\omega)) & = & \omega \end{array} \qquad \begin{array}{ccc} * : \Lambda^2 & \to & \Lambda^2 \\ dx^1 \wedge dx^2 & \mapsto & dx^3 \wedge dx^4 \\ dx^2 \wedge dx^3 & \mapsto & dx^1 \wedge dx^4 \\ dx^3 \wedge dx^1 & \mapsto & dx^2 \wedge dx^4 \\ dx^1 \wedge dx^4 & \mapsto & -dx^2 \wedge dx^3 \\ dx^2 \wedge dx^4 & \mapsto & -dx^3 \wedge dx^1 \\ dx^3 \wedge dx^4 & \mapsto & -dx^1 \wedge dx^2 \\ *(*(\omega)) & = & -\omega \end{array}$$

Note in particular that the Hodge-$*$is a complex structure on $\Lambda^2$. Next we compute $d(*F)$.

$$F = \sum_i E_i dx^i \wedge dx^4 + \sum_{(i,j,k)} B_i dx^j \wedge dx^k$$
$$*F = -\sum_{(i,j,k)} E_i dx^j \wedge dx^k + \sum_i B_i dx^i \wedge dx^4$$

The computation for $d(*F)$ is therefore the same as for $dF$ (using $df = \sum \frac{\partial}{\partial x}^i dx^i$), we get:

$$d(*F) = -\mathrm{div}\,\vec{E}\,dx^1 \wedge dx^2 \wedge dx^3 + \sum_{(i,j,k)} \big(\mathrm{rot}\,\vec{B} - \frac{\partial}{\partial x^4}\vec{E}\,\big)_i dx^j \wedge dx^k \wedge dx^4.$$

This is another place where a physicist would keep the factors $c$. Rewrite the last line, using $dx_4 = c\,dt$ and $\epsilon_0 c^2 = 1/\mu_0$:

$$\epsilon_0 d(*F) = -\mathrm{div}\,\vec{D}\,dx^1 \wedge dx^2 \wedge dx^3 + \frac{1}{c}\sum_{(i,j,k)} \big(\mathrm{rot}\,\vec{H} - \frac{\partial}{\partial t}\vec{D}\,\big)_i dx^j \wedge dx^k \wedge dt.$$

The 4-vector with components *current density* and *charge density* is called **4-current** $J$:

$$J := \begin{pmatrix} \mathrm{rot}\,\vec{H} - \frac{\partial}{\partial t}\vec{D} \\ \mathrm{div}\,\vec{D} \end{pmatrix} = \begin{pmatrix} \vec{j} \\ \rho \end{pmatrix} = \begin{pmatrix} \text{current density} \\ \text{charge density} \end{pmatrix}.$$

We denote the vector that corresponds to a 1-form $\omega \in \Lambda^1$ by $\omega^\#$. Then we can rewrite the non-homogenous Maxwell equations as

$$(*d*\;F)^\# = J.$$

One checks easily that the first two examples satisfy this equation away from the charges. — Since the Hodge-$*$is invariant under Lorentz transformations ($= \langle\langle.,.\rangle\rangle$-isometries), we have proved that the Maxwell equations are Lorentz invariant. The operators $d$ and $*$

exist tangent space wise on any Lorentz manifold. Therefore it is clear how to generalize the Maxwell equations to Lorentz manifolds. Experimental evidence about electromagnetic waves in curved space can only come from astronomical observations. So far there are no doubts that $dF = 0, (\ast d \ast F)^{\#} = J$ are the correct equations. If we agree to make this generalization of what we call the Maxwell equations then we have also proved:

THEOREM. The Maxwell equations are conformally invariant.
This means: If we change the metric conformally then solutions of Maxwell's equations remain solutions.

**Plane Waves**

With a function $f$ of period 1 consider the fields

$$\vec{E} = f(\omega \cdot (x^1 - t)) \cdot \vec{e}_2, \quad \vec{B} = f(\omega \cdot (x^1 - t)) \cdot \vec{e}_3,$$

With an antiderivative $g,\ g' = f$ define $h(x^1, t) := g(\omega \cdot (x^1 - t))$. We can write the corresponding Faraday form $F$ and $\ast F$ as

$$F = f(\omega \cdot (x^1 - t)) \cdot (dx^2 \wedge dt + dx^1 \wedge dx^2) = \frac{1}{\omega} dh \wedge dx^2$$
$$\ast F = f(\omega \cdot (x^1 - t)) \cdot (-dx^3 \wedge dx^1 + dx^3 \wedge dt) = \frac{1}{\omega} dh \wedge dx^3$$

Clearly, $dF = 0, d\ast F = 0$, so that the homogenous Maxwell equations are satisfied. We interpret this solution as a wave, traveling in the $x^1$-direction. What does another observer, traveling with velocity $v$ in the $x^1$-direction, see? His rest frame is

$$\vec{f}_1 = (\vec{e}_1 + v\vec{e}_4)/\sqrt{1 - v^2}, \quad \vec{f}_2 = \vec{e}_2, \quad \vec{f}_3 = \vec{e}_3, \quad \vec{f}_4 = (v\vec{e}_1 + \vec{e}_4)/\sqrt{1 - v^2},$$

and his coordinates $(y^1, y^2, y^3, \tau)$ are

$$y^1 = (x^1 - vt)/\sqrt{1 - v^2},\ y^2 = x^2,\ y^3 = x^3,\ \tau = (-vx^1 + t)/\sqrt{1 - v^2}.$$

His electric field $\vec{E}_{new}$ is given by

$$(\vec{E}_{new})_2 = F(\vec{f}_2, \vec{f}_4) = f(\omega(x^1 - t)) \cdot \left(\frac{1}{\sqrt{1 - v^2}} - \frac{v}{\sqrt{1 - v^2}}\right) = f(\omega(x^1 - t)) \cdot \sqrt{\frac{1 - v}{1 + v}}$$
$$= f(\omega \cdot \sqrt{\frac{1 - v}{1 + v}}(y^1 - \tau)) \cdot \sqrt{\frac{1 - v}{1 + v}},$$
$$f_{new} = f \cdot \sqrt{\frac{1 - v}{1 + v}}, \qquad \omega_{new} = \omega \cdot \sqrt{\frac{1 - v}{1 + v}}.$$

This result says that the observer that moves away with velocity $v$ sees frequency and amplitude smaller by the factor $\sqrt{(1 - v)/(1 + v)}$. For the frequency this is in agreement with our earlier discussion of the Doppler ratio in which we used light rays rather than waves. The weakening of the amplitude is also important in astronomy, it is called *Ausdünning* (thinning?).

**Charged Particles**

Assume the 2-form $F$ satisfies Maxwell's homogenous equations, $dF = 0,\ d{*}F = 0$. Define a skewsymmetric endomorphism field $\tilde{F}$ by

$$F(X,Z) =: \langle\langle X, \widehat{F}(Z)\rangle\rangle = -\langle\langle Z, \widehat{F}(X)\rangle\rangle.$$

With an orthogonal basis $\{\vec{e}_1, \ldots, \vec{e}_4\}$ we have

$$\widehat{F}(Z) = \sum_i \frac{\langle\langle \widehat{F}(Z), \vec{e}_i\rangle\rangle}{\langle\langle \vec{e}_i, \vec{e}_i\rangle\rangle}\vec{e}_i = \sum_i \frac{F(\vec{e}_i, Z)}{\langle\langle \vec{e}_i, \vec{e}_i\rangle\rangle}\vec{e}_i.$$

The world line $c(t)$ of a particle with charge $e$ and rest mass $m$ satisfies the ODE

$$\ddot{c}(t) = \frac{e}{m}\widehat{F}(\dot{c}(t)) \qquad \text{(so called Lorentz Force)}$$

Such a world line is automatically parametrized proportional to proper time:

$$\frac{d}{dt}\langle\langle \dot{c}(t), \dot{c}(t)\rangle\rangle = 2\langle\langle \ddot{c}, \dot{c}\rangle\rangle = \frac{2e}{m}\langle\langle \widehat{F}(\dot{c}(t)),\ \dot{c}(t)\rangle\rangle = 0.$$

Usually one therefore assumes $\langle\langle \dot{c}, \dot{c}\rangle\rangle|_{t=0} = -1$, so that for each $t$ the rest frame of the particle has $\vec{e}_4 = \dot{c}(t)$. This implies

$$\widehat{F}(\dot{c}(t)) = \sum_i \frac{F(\vec{e}_i, \vec{e}_4)}{\langle\langle \vec{e}_i, \vec{e}_i\rangle\rangle}\vec{e}_i = \vec{E}\,.$$

We see that in the *rest frame* of the particle only the electric field exerts a force, the magnetic field may be nonzero but it does not contribute to the acceleration of the world line.

Next assume that in the rest frame of an inertial observer we have a field

$$\vec{E} = 0,\ \vec{B} = (0,0,b), \quad \text{hence} \quad F = b \cdot dx^1 \wedge dx^2.$$

The charged particle may have a world line $c(t)$ whose tangent field can be written as

$$\dot{c}(t) = (v(t)\cos\alpha(t), v(t)\sin\alpha(t), 0, 1)/\sqrt{1-v^2}.$$

Since $F(\vec{e}_1, \vec{e}_2) = b$ is the only non-vanishing component, we find

$$e\widehat{F}(\dot{c}) = \widehat{F}(\cos\alpha\vec{e}_1 + \sin\alpha\vec{e}_2)\frac{ev}{\sqrt{1-v^2}} = \frac{-b \cdot ev}{\sqrt{1-v^2}}(\vec{e}_2\cos\alpha - \vec{e}_1\sin\alpha).$$

Compare with the 3-dimensional formulation of the Lorentz force:

$$force = (charge \cdot velocity) \times magneticField.$$

We have computed the force along a given world line, now we want to solve the corresponding ODE. In the rest system of the above observer (the one with the constant magnetic field and a world line $t \to (0,0,0,t)$ ) we look for a circular orbit (with its proper time $s$ as parameter):

$$\begin{aligned}
c(s) &:= (r\cos\omega s, r\sin\omega s, 0, \tau s)\\
c'(s) &:= (-r\omega\sin\omega s, r\omega\cos\omega s, 0, \tau) = \frac{1}{\sqrt{1-v^2}}(v_1, v_2, 0, 1),\ \tau = \frac{1}{\sqrt{1-v^2}},\ v = \frac{r\omega}{\tau},\\
c''(s) &= -r\omega^2(\cos\omega s, \sin\omega s, 0, 0) = \frac{e}{m}\widehat{F}(c') = \frac{eb}{m} r\omega(\cos\omega s, \sin\omega s, 0, 0),\quad \omega = \frac{-eb}{m}.
\end{aligned}$$

What is the interpretation of this computation? The non-inertial orbiting particle has $s$ as proper time. It observes $\omega/2\pi$ as its rotation frequency. The parameter $m$ is its rest mass. The force that causes $c''$ to be nonzero is purely Coulomb, i.e. $e\vec{E}$. — The other observer has an inertial laboratory whose points have world lines $t \to (x, y, z, t)$, where $t$ is the synchronized proper time on all these laboratory world lines. The rotating particle meets the world line $(r, 0, 0, t)$ at $s_0 = 0, s_1 = 1/\omega, \ldots$, that is at laboratory times $t_0 = 0, t_1 = \tau \cdot s_1 = \tau/\omega$, hence $\omega_{lab} = \omega/\tau$. In other words, one complete rotation takes the factor $\tau$ more time in the laboratory system, in agreement with the more elementary discussion in the first lecture. If a collision were observed the particle mass would be larger than the rest mass, $m_{lab} = \tau \cdot m$. And, there are no electric fields observed in the laboratory, only the constant magnetic field. – Note $\omega \cdot m = \omega_{lab} \cdot m_{lab}$.

## Aberration of Light

We need to understand in which way the sky is different when we look at it from opposite points of the orbit of the earth around the sun (the orbit velocity is 30 km/sec $= 10^{-4}c$). We can think of the sky as the 2-sphere at infinity, as the 2-sphere of directions. This 2-sphere is the (for example: unit) 2-sphere in the observer's simultaneity space $e_4^\perp$. How does one translate the statement: *We see a star in direction* $\vec{e} \in \mathbb{S}^2 \subset e_4^\perp$ into the 4-dim language of Special Relativity? Clearly, the world lines of the incoming light signals are straight lines on the observers backwards light cone. And the question means: *which such null world line corresponds to the 3-dim direction* $\vec{e}$*?* We can translate the simulaneity space with its unit 2-sphere one unit of time backwards and observe that this translated space intersects the light cone in the unit 2-sphere. Therefore we identify the 3-dim direction $\vec{e}$ with the null ray $(\vec{e}, 1) \cdot \mathbb{R}$. Since these null world lines of light from the stars are independent of the observers we need to explicitly say, in a Lorentz invariant formulation, how we obtain the angle $\varphi$ with $\cos\varphi = \langle \vec{e}, \vec{f} \rangle$ between two directions $\vec{e}, \vec{f}$ of the above observer from the null world lines:

Let $\tilde{e}_4 = (\vec{v}, 1)/\sqrt{1-|v|^2}$ be the time unit vector of another observer. Intersection of the light cone with $\tilde{e}_4$'s simultaneity space at time $-1$ means: choose the tangent vectors $\dot{c}$ to the null world lines $c$ such that $\langle\langle \dot{c}, \tilde{e}_4 \rangle\rangle = -1$. This implies for the difference $\dot{c} - \tilde{e}_4 \in \tilde{e}_4^\perp$, in other words, these difference vectors are the unit vectors pointing to the stars in the rest space of the second observer. Therefore we have for the angle between two stars as seen by the second observer

$$\cos\tilde{\varphi} = \langle\langle \dot{c}_1 - \tilde{e}_4, \dot{c}_2 - \tilde{e}_4 \rangle\rangle.$$

We do this computation for the directions $\vec{e}, \vec{f}$ of the first observer. First we have the reparametrization factor $\lambda(\vec{e}, \vec{v})$ so that

$$\langle\langle \lambda(\vec{e}, \vec{v}) \cdot (\vec{e}, 1),\ \frac{(\vec{v}, 1)}{\sqrt{1 - |v|^2}} \rangle\rangle = -1 \quad \text{or} \quad \lambda(\vec{e}, \vec{v}) = \frac{-\sqrt{1 - |v|^2}}{(\langle \vec{e}, \vec{v} \rangle - 1)}$$

The new direction vectors in the second rest space are the differences

$$\vec{\tilde{e}} := \lambda(\vec{e}, \vec{v}) \cdot (\vec{e}, 1) - \frac{(\vec{v}, 1)}{\sqrt{1 - |v|^2}}, \qquad \vec{\tilde{f}} := \lambda(\vec{f}, \vec{v}) \cdot (\vec{f}, 1) - \frac{(\vec{v}, 1)}{\sqrt{1 - |v|^2}}.$$

Their $\langle\langle ., . \rangle\rangle$-scalar product is the Euclidean scalar product in the rest space and therefore gives $\cos\tilde{\varphi}$.
Since the orbit speed of the earth is only $10^{-4}c$ we simplify these formulas by dropping all terms that are at least quadratic in $\vec{v}$. This gives the

small $|\vec{v}|$ approximation of the aberration of light:

$$\lambda(\vec{e}, \vec{v}) = 1 + \langle \vec{e}, \vec{v} \rangle, \quad \vec{\tilde{e}} = \big(\vec{e} - \vec{v} + \langle \vec{e}, \vec{v} \rangle \vec{e},\ \langle \vec{e}, \vec{v} \rangle\big), \quad \vec{\tilde{f}} = \big(\vec{f} - \vec{v} + \langle \vec{f}, \vec{v} \rangle \vec{f},\ \langle \vec{f}, \vec{v} \rangle\big)$$

$$\cos\tilde{\varphi} \approx \langle\langle \vec{\tilde{e}}, \vec{\tilde{f}} \rangle\rangle \approx \cos\varphi - (\langle \vec{e}, \vec{v} \rangle + \langle \vec{f}, \vec{v} \rangle) \cdot (1 - \cos\varphi) + o(|\vec{v}|^2).$$

**Images to sort out**

It needs some practice to translate 3-dim descriptions into Special Relativity. For example: An observer cannot move from one point of its rest space to another point, as soon as he moves, his rest space changes. He can only receive light signals and know later what happened in his rest space. To discuss relative movement of two systems the simplest situation is: pick the world lines of two observers in relative motion, draw a segment in the rest spaces of each observer, consider the world lines of all the points of the two segments in the two systems. At the instant where a world line $c_1$ of a segment point $a_1$ in the first system intersects the world line $c_2$ of a segment point $b_2$ of the second system, at that instant the first observer says: *the other guy's point b just now flies past my point a* and the second observer says: *the other guy's point a just now flies past my point b.*
Starting from this picture one can answer questions like: How long does it take, in each system, for one point of a segment to fly past the other segment? There is no problem with the length of a segment in its own system, but what does the other observer take as the segment length?

# Lightcone Geometry

Null Geodesics, Distance Measurements, Red Shift

Remark. At some point I should present a homogenous, flat (i.e. curvature tensor $R = 0$) example with incomplete null geodesics.

In this section $M, g$ is a Lorentz manifold (i.e. $g$ has the signature of Special relativity). We discuss properties of null geodesics, emphasizing similarities with Special Relativity.

1. Geodesic Lightcones. The null vectors $v \in T_pM$ form the light cone in $T_pM$. Initial data $\{p, v\}$ for geodesics integrate to null geodesics because

$$D_{\gamma'}\gamma' = 0 \text{ and } g(\gamma'(0), \gamma'(0)) = 0 \Rightarrow g(\gamma'(t), \gamma'(t)) = 0.$$

We say *these null geodesics form the geodesic light cone with vertex* $p$. This lightcone is the image of the light cone in $T_pM$ under the exponential map.

2. Tangential Jacobi Fields. Let $\gamma_\epsilon(s)$ be a family of null geodesics with $\gamma_\epsilon(0) = p$. The Jacobi fields $J(s) := \frac{\partial}{\partial\epsilon}\gamma_\epsilon(s)|_{\epsilon=0}$ are tangent vectors to the light cone. From $g(\gamma'_\epsilon(s), \gamma'_\epsilon(s)) = 0$ we obtain by differentiation

$$0 = \frac{\partial}{\partial\epsilon} g(\gamma'_\epsilon(s), \gamma'_\epsilon(s))|_{\epsilon=0} = 2g(\gamma'(s), \frac{D}{ds}J(s))$$

The derivatives of such Jacobi fields therefore lie in the subspace $T := \{u : g(\gamma'(s), u) = 0\}$. In Special Relativity this equation defines the tangent spaces of the light cone. Because

$$0 = g(\gamma'(s), \frac{D}{ds}J(s)) = \frac{d}{ds}g(\gamma'(s), J(s)) \Rightarrow g(\gamma'(s), J(s)) = const = g(\gamma'(0), J(0)) = 0.$$

we also have in $M, g$ that the tangent vectors $J(s)$ to the lightcone satisfy this equation.

3. Parallel Tangent Planes. In Special Relativity the tangent space to the light cone at one point of a light ray is tangential to the light cone along the whole ray. Similarly in general: Let $u(s_0)$ be a tangent vector to the light cone at the point $\gamma(s_0)$ of the null geodesic $\gamma(.)$. Extend this tangent vector to a *parallel* field $u(.)$ along $\gamma(.)$.

Claim: $u(s)$ is tangent to the light cone at $\gamma(s)$.

Proof: $\frac{D}{ds}u = 0 \Rightarrow g(\gamma'(s), u(s)) = const = g(\gamma'(s_0), u(s_0)) = 0$.

This says that the geodesic light cone geometry is as well behaved as in Special Relativity: We can pick **any** point on a null geodesic and consider the light cone with that vertex. We proved that all these light cones are tangential to each other and their common tangent planes are parallel along the common null geodesic. — In Riemannian geometry often problems are caused by the rotation of Jacobi fields. Along the light cone we are in a better situation.

4. Parallel Null Vectors. If one picks a null vector $u$ that is linearly independent of $\gamma'(s_0)$ then it is tangential only to the light cone with vertex $\gamma(s_0)$ and otherwise transversal

to the common tangent plane of the light cones containing $\gamma(.)$. Of course, if we parallel translate $u$ along $\gamma$ then it stays a null vector because $g(u(s), u(s))$ is constant.

5. Non-Tangential Jacobi Fields. We consider a family $\gamma_\epsilon(s)$ of null geodesics that are **not** assumed to have a common initial point. As in that special case we obtain for the Jacobi field $J(s) := \frac{\partial}{\partial\epsilon}\gamma_\epsilon(s)|_{\epsilon=0}$ that $g(\frac{D}{ds}J(s), \gamma'(s)) = 0$. This says that the *derivative* of $J$ lies in the tangent spaces of the light cone along $\gamma$. We can therefore decompose

$$J(s) = u(s) + T(s),$$

where $u(.)$ is **parallel** along $\gamma$ and $T(.)$ is tangential to the light cone.

Warning: both, $u(.), T(.)$ are in general **not** Jacobi fields. Proof of the decomposition: begin with $J(0) = u(0) + T(0)$ as desired, extend $u(0)$ to a parallel field $u(.)$ and define $T(s) := J(s) - u(s)$. Then $\frac{D}{ds}T(s) = \frac{D}{ds}J(s)$, hence $\frac{d}{ds}g(\gamma'(s), T(s)) = 0$ and $g(\gamma'(s), T(s)) = const = g(\gamma'(0), T(0)) = 0$, as claimed.

6. Killing Fields. A map $A : M \to M$ is called a Pseudo-Riemannian isometry (often Lorentz isometry for short) if it satisfies for arbitrary tangent vectors $Y, Z$

$$g(TA \cdot Y, TA \cdot Z) = g(Y, Z).$$

Let $A_t$ be a family of Lorentz isometries with $A_0 = \mathrm{id}$.

Definition of a **Killing field** $\quad X(p) := \frac{\partial}{\partial t}A_t(p)|_{t=0}.$

Claim: The covariant differential $DX$ of a Killing field is a skew-symmetric endomorphism field, i.e. $g(D_Y X, Y) = 0$. And vice versa, the flow of a vector field $X$ with skew-symmetric $DX$ consists of Lorentz isometries, i.e., $X$ is a Killing field.

Choose a curve $p(s),\ p(0) = p,\ p'(0) = Y$. Then

$$\begin{aligned}
0 = \frac{d}{dt}g(TA_t \cdot Y, TA_t \cdot Y)|_{t=0} &= 2g(\frac{D}{dt}\big(\frac{d}{ds}A_t(p(s))|_{s=0}\big), TA_t \cdot Y)|_{t=0} \\
&= 2g(\frac{D}{ds}\big(\frac{d}{dt}A_t(p(s))|_{t=0}\big)|_{s=0},\ Y) \\
&= 2g(\frac{D}{ds}\big(X(p(s)\big)|_{s=0},\ Y) \\
0 = \qquad\qquad &= 2g(D_Y X(p), Y).
\end{aligned}$$

In the reverse direction, if the covariant differential of a vector field $X$ is a skew-symmetric endomorphism field and if $A_t$ denotes the flow of this vector field then the above computation (read backwards) shows that the scalar products $g(TA_t \cdot Y, TA_t \cdot Y)$ do not depend on $t$ so that the $A_t$ are indeed Lorentz isometries.

7. Killing Fields and Conserved Quantities. If $X$ is a Killing field and $\gamma$ is a geodesic (for example a force free world line) then

$$g(X(\gamma(s)), \gamma'(s)) = const$$

Proof: $\frac{d}{ds} g(X(\gamma(s)), \gamma'(s)) = g(\frac{D}{ds} X(\gamma(s)), \gamma'(s)) + g(X(\gamma(s)), \frac{D}{ds}\gamma'(s))$, the first term is zero by skew-symmetry of $DX$, the second because of the geodesic equation.

We will have to discuss the question: why do we observe conserved quantities even though our cosmological Lorentz manifold has no Killing fields. The following will be needed.

8. Second Order PDE for Killing Fields. The differential equation for Killing fields – the skew-symmetry of $DX$ – cannot be solved on arbitrary manifolds, because there exist not enough Isometries. We want to connect this more directly with the equation and we will find that the Killing fields also solve a second order PDE that *reduces along geodesics to second order ODEs*. This shows that the initial value and the initial derivative at one point, i.e. $X(p), DX|_p$, determine a Killing field uniquely – while not every such candidate really is a Killing field.
The flow of a Killing field moves each geodesic through a family of geodesics, so that the restriction of a Killing field to a geodesic is a Jacobi field. That means, for every tangent vector of a geodesic, $\gamma'$, the Killing field $X$ has to satisfy $D^2_{\gamma',\gamma'} X + R(X, \gamma')\gamma' = 0$. Hence we have for all tangent vectors $Y, Z$

$$D^2_{Y+Z,Y+Z} X + R(X, Y + Z)Y + Z = 0,$$

$$D^2_{Y,Y} X + R(X, Y)Y = 0, \quad D^2_{Z,Z} X + R(X, Z)Z = 0,$$

by subtraction $\quad D^2_{Y,Z} X + D^2_{Z,Y} X + R(X, Y)Z + R(X, Z)Y = 0,$

by definition $\quad D^2_{Y,Z} X - D^2_{Z,Y} X = R(Y, Z)X.$

Finally $\quad 2D^2_{Y,Z} X + R(X, Y)Z = R(Z, X)Y + R(Y, Z)X,$

with 1.Bianchi $\quad D^2_{Y,Z} X + R(X, Y)Z = 0.$

Remark. If one tries to *construct* $X$ by solving Jacobi equations along radial geodesics, then one can guarantee the correct second derivative of $X$ only in the direction of those geodesics – while the second order PDE (derived above) requires much more.

The most important class of cosmological models will be conformally flat. Recall also that the Maxwell equations are conformally invariant. Conformal changes of the metric will therefore be important for the discussion of such cosmological models. We will see that null geodesics remain null geodesics under conformal changes – except that they are no longer parametrized so that the tangent field is parallel. We briefly discuss such ("non affine") parametrizations of geodesics.

9. Geodesics with non affine parametrization. Start with $\frac{D}{ds} c'(s) = 0$ and reparametrize $s = \varphi(\sigma)$, i.e., consider $\gamma(\sigma) := c(\varphi(\sigma))$.

$$
\begin{aligned}
\frac{d}{d\sigma}\gamma(\sigma) &= c'(\varphi(\sigma)) \cdot \frac{d}{d\sigma}\varphi(\sigma) \\
\frac{D}{d\sigma}\Big(\frac{d}{d\sigma}\gamma(\sigma)\Big) &= \frac{D}{ds}c'|_{\varphi(\sigma)} \cdot (\frac{d}{d\sigma}\varphi(\sigma))^2 + c'(\varphi(\sigma)) \cdot \varphi''(\sigma) \\
&= \frac{\varphi''}{\varphi'}(\sigma) \cdot \frac{d}{d\sigma}\gamma(\sigma).
\end{aligned}
$$

Vice versa, consider a curve such that the covariant derivative of the tangent field is proportional to the tangent field:

$$
\frac{D}{d\sigma}\Big(\frac{d}{d\sigma}\gamma(\sigma)\Big) = m(\sigma)\frac{d}{d\sigma}\gamma(\sigma).
$$

We insert a parameter change $\sigma = \psi(s)$ and determine $\psi(.)$ so that $\frac{D}{ds}(\frac{d}{ds}(\gamma(\psi(s)))) = 0$:

$$
\begin{aligned}
\frac{D}{ds}\Big(\frac{d}{ds}(\gamma(\psi(s)))\Big) &= \frac{D}{ds}\Big(\psi'(s) \cdot \frac{d}{d\sigma}\gamma(\sigma)))\Big) \\
&= \psi''(s) \cdot \frac{d}{d\sigma}\gamma(\sigma)) + (\psi'(s))^2 \cdot \frac{D}{d\sigma}(\frac{d}{d\sigma}\gamma(\sigma)) \\
&= \big(\psi''(s) + \psi'(s)^2 m(\sigma)\big)\frac{d}{d\sigma}\gamma(\sigma).
\end{aligned}
$$

Indeed, by solving the ODE

$$
\psi''(s) = -m(\psi(s)) \cdot \psi'(s)^2,
$$

we can return to an affine parametrization of the given curve. In our application the function $m(.)$ will be related to the conformal factor and we can explicitly solve the ODE. Here we only remark: **if** $m = \varphi''/\varphi'$ then, expectedly, the ODE is solved by $\psi := \varphi^{-1}$.

## 10. Null Geodesics under Conformal Changes of $g(.,.)$.

We consider the conformally changed metric $\tilde{g}(X,Y) := \lambda^{-2}g(X,Y),\ \lambda > 0$ (the exponent $-2$ turns out to be slightly more convenient). We define the difference tensor between the two covariant derivatives and find from a short computation

$$
\Gamma(X,Y) := \tilde{D}_X Y - D_X Y = -\frac{T_Y\lambda}{\lambda}X - \frac{T_X\lambda}{\lambda}Y + g(X,Y)\mathrm{grad}\,\lambda
$$

Remark. From $g(\mathrm{grad}\,\lambda, X) = T_X\lambda = \tilde{g}(\widetilde{\mathrm{grad}}\,\lambda, X) = \lambda^{-2}g(\widetilde{\mathrm{grad}}\,\lambda, X)$ we see

$$
\mathrm{grad}\,\lambda = \lambda^{-2}\widetilde{\mathrm{grad}}\,\lambda, \qquad g(X,Y)\mathrm{grad}\,\lambda = \tilde{g}(X,Y)\widetilde{\mathrm{grad}}\,\lambda.
$$

In Riemannian geometry geodesics do not remain geodesics under conformal changes of the metric because of the Term $g(X,Y)\mathrm{grad}\,\lambda$ in the difference tensor $\Gamma$. This term drops out for null geodesics $\gamma$ with $\frac{D}{d\sigma}\gamma'(\sigma) = 0,\ g(\gamma'(\sigma),\gamma'(\sigma)) = 0$. We obtain

$$\frac{D}{d\sigma}\gamma'(\sigma) = 0 \quad \Rightarrow \quad \frac{\tilde{D}}{d\sigma}\gamma'(\sigma) = \Gamma(\gamma', \gamma')(\sigma) = -2\frac{T_{\gamma'}\lambda}{\lambda} \cdot \frac{d}{d\sigma}\gamma(\sigma) = -2\frac{\frac{d}{d\sigma}\lambda(\gamma(\sigma))}{\lambda} \cdot \frac{d}{d\sigma}\gamma(\sigma).$$

So indeed, $\gamma$ is also a null geodesic for $\tilde{g}(.,.)$, but not with affine parametrization. The function $m(\sigma)$ in the above reparametrization computation equals in the present case $m(\sigma) = -2\frac{d}{d\sigma}\lambda(\gamma(\sigma))/\lambda$, so that the second order ODE $\psi''(s) = -m(\psi(s))\psi'(s)^2$ can be integrated once to:

$$\psi'(s) = \lambda^2(\gamma(\psi(s))), \qquad \text{since this solves the ODE:}$$
$$\psi''(s) = 2\lambda \cdot \frac{d}{d\sigma}\lambda(\gamma(\sigma)) \cdot \psi'(s) \cdot \frac{\psi'}{\lambda^2} = -m(\psi(s))\psi'(s)^2.$$

11. Reshift in Lorentz Manifolds. Consider the following experiment:
A source S emits light signals of known frequency. These signals are observed by an observer O. Source and observer have their world lines parametrized by proper time. The unit tangent vectors of these world lines are denoted $u$ (source) and $v$ (observer), $g(u, u) = -1 = g(v, v)$. The light signals travel on null geodesics that join the world lines of S and O. Note that the light signals emitted at source time $t$ travel on all the null geodesics of the light cone that has its vertex at the point of emission. These light rays thus fill a 3-dimensional surface and if the world line of O meets the light cone for emission time $t$, then, by transversality, O's world line continues to meet the light cones from later emission points of S. So we have a family of null geodesics that join the two world lines. We assume affine parametrizations for these null geodesics on the interval $[0, 1]$.
Now we differentiate the family and have a Jacobi field $J$ along one null geodesic $c$ from S to O. $J(0)$ is proportional to $u$, we may assume $J(0) = u$ (for example with the interpretation: the time signals are sent at unit time intervals). $J(1)$ is proportional to $v$ and the factor determines how the observed frequency differs from the emitted frequency.
Standard conventions are that the ratio of the source frequency divided by the observer frequency is written as $1+z$ and this ratio (or also the number $z$) is called **red shift** because for most astronomical situations the source frequency is larger and red shift is meant to indicate a change towards slower frequencies because *red* is at the slow end of the visible light spectrum.

Definition of red shift: $$\frac{\omega_S}{\omega_O} =: 1 + z.$$

Claim: $$\frac{\omega_S}{\omega_O} = \frac{g(u, c')}{g(v, c')}.$$

Note that an affine change of parameter on the null geodesic $c$ drops out of this quotient. Note also that this formula captures red shift as a natural geometric quantity. We do not have distances associated to an indefinite metric, but in addition to proper time on world lines we now also have red shift as a geometric quantity. And in the development of astronomy the discovery of red shift was certainly a dramatic moment. Before I prove the claim first an

EXAMPLE FROM SPECIAL RELATIVITY.
Take $u := (0,0,0,1)$ and $v = (a,0,0,1)/\sqrt{1-a^2}$ and a null ray $c$ in the $x$-direction, $c' = (1,0,0,1)$. Then we have

$$g(u,c') = -1,\ g(v,c') = -\sqrt{\frac{(1-a)}{(1+a)}} \quad \Rightarrow \quad \frac{g(u,c')}{g(v,c')} = \sqrt{\frac{(1+a)}{(1-a)}}.$$

This is the same answer that we obtained before: for source and observer flying apart the emitted frequency is larger by the given factor than the observed frequency.

PROOF OF THE CLAIM. We arranged the variation by null geodesics so that $J(0) = u$ and $J(1)$ is proportional to $v$, say $J(1) = \mu v$. Observe that $\mu$ is the quantity we need to determine because time signals that are emitted one unit of time apart are received $\mu$ units of time apart. Rephrased as frequency change this says: $\omega_S/\omega_O = \mu$. We need point 5. above:

$$g(\mu v, c'(1)) = g(J(1), c'(1)) \overset{(5.)}{=} g(J(0), c'(0)) = g(u, c'(0)) \quad \Rightarrow \quad \frac{\omega_S}{\omega_O} = \mu = \frac{g(u,c'(0))}{g(v,c'(1))}.$$

12. RESHIFT AND CONFORMAL CHANGE OF $g(.,.)$. We want to add a conformal change of the metric, $\tilde{g} = \lambda^{-2} g$ to the previous red shift discussion. This ought to be possible since the null geodesics that connect the world lines of source and observer remain null geodesics, we only have to correct for the change of paramerization. We call the $g$-geodesics $\gamma(\sigma)$ and the $\tilde{g}$-geodesics $c(s) = \gamma(\psi(s))$ with $\psi'(s) = \lambda^2(\gamma(\psi(s)))$ and of course $c'(s) = \frac{d}{d\sigma}\gamma(\sigma) \cdot \psi'(s) = \lambda^2 \gamma'(\sigma)$. The timelike unit vectors of S and O are easily changed $\tilde{u} = \lambda(S) u,\ \tilde{v} = \lambda(O) v$ so that again $\tilde{g}(\tilde{u},\tilde{u}) = -1 = \tilde{g}(\tilde{v},\tilde{v})$. If we forget to take the parametrization change of the null geodesics into account this would indicate that the frequencies are changed with $\lambda^{-1}$. This statement by itsself does not make sense since there are no experiments that connect frequencies in different Lorentz manifolds. The correct computation of the frequency ratio takes the reparametrization of the null geodesics into account:

$$\frac{\tilde{\omega}_S}{\tilde{\omega}_O} = \frac{\tilde{g}(\tilde{u}, c'|_S)}{\tilde{g}(\tilde{v}, c'|_O)} = \frac{\lambda^{-2}(S) \cdot g(\lambda(S)u, \lambda^2(S)\gamma'|_S)}{\lambda^{-2}(O) \cdot g(\lambda(O)v, \lambda^2(O)\gamma'|_O)} = \frac{\omega_S}{\omega_O} \cdot \frac{\lambda(S)}{\lambda(O)}$$

In our discussion of cosmological observations it will be very convenient that the change of red shift under conformal changes of the metric does not depend on solutions of ODEs but involves only the ratio of the conformal factors at the source S and the observer O.

13. QUOTIENT GEOMETRY ON THE LIGHT CONE. The induced metric on any light cone $LC$ is degenerate: let $c(s)$ be a null geodesic on $LC$, then we have for **all** $v \in T_{c(s)}LC$ that $g(c'(s), v) = 0$. It is therefore useful to introduce the quotient geometry by defining

equivalence classes: $\qquad [v] := v + \mathbb{R}c'(s).$

We have a well defined positive definite scalar product on the quotient of $T_{c(s)}LC$:

$$\begin{aligned} g([v],[w]) &:= g(v+\lambda c', w+\mu c') \\ &= g(v,w) + \lambda g(c',w) + \mu g(v,c') + \lambda\mu g(c',c') = g(v,w). \end{aligned}$$

We have a covariant derivative on the quotient bundle along $c(s)$:
let $[v](s) = [v(s) + \lambda(s)c'(s)]$ then we can define (with $(\frac{D}{ds}v(s))^{tang}$ denoting the $LC$-tangential component of $\frac{D}{ds}v(s)$ )

$$\frac{D}{ds}[v](s) := [(\frac{D}{ds}v(s))^{tang} + \lambda'(s)c'(s)] = [(\frac{D}{ds}v(s))^{tang}] \quad \text{and hence have}$$

$$\begin{aligned} \frac{d}{ds}g([v](s),[w](s)) &= \frac{d}{ds}g(v(s),w(s)) = g(\frac{D}{ds}v(s),w(s)) + g(v(s),\frac{D}{ds}w(s)) \\ &= g(\frac{D}{ds}[v](s),[w](s)) + g([v](s),\frac{D}{ds}[w](s)). \end{aligned}$$

Also, the Jacobi equation descends to the quotient.
Since $R(v+\lambda c', c')c' = R(v,c')c' \in T_{c(s)}TC$ we can define

$$[R]([v],c')c' := [R(v+\lambda c', c')c'],$$

so that $[R]$ is a symmetric operator on the 2-dimensional quotient space at $c(s)$. Finally, since a tangential Jacobi field $J(s)$ has a tangential covariant derivative, we have the twodimensional quotient equation

$$\frac{D}{ds}(\frac{D}{ds}[J(s)] + [R]([J](s),c')c' = 0.$$

The two eigenvalues of $[R]$ will be important for measurement discussions.

## Distance Measurements in Relativity Theory

First we describe in Newtonian language methods to determine distances that are used in astronomy.

1. Luminosity Comparison. Because the area of spheres grows as $r^2$ (with $r$ the distance) the intensity of light goes down as $r^{-2}$. Since the Fraunhofer spectra of the stars are so complicated one is often in a position to say that two stars have the **same** absolute brightness. Then the ratio of their apparent brightnesses in the sky is the square of the ratio of their distances (twice as far away = one quarter as bright). This method is applied not only to individual stars in our and in nearby galaxies but also to whole galaxies or other objects (e.g. certain supernovae) where one has a good argument why their absolute brightness should be the same. – For each type of object of known absolute brightness one needs at least one object of known absolute distance, otherwise the method cannot be applied to that type of object.

2. Trigonometric Distance via Parallax. The fact, that a triangle is determined if one knows one edge and two angles, allows to find the position of a boat if it can see two known light houses. The same principle is used in astronomy by observing a star from two opposite points of the orbit of the earth around the sun. The main observational problem is to compare the two directions in which one sees the star from those two opposite positions. Today one has an enormous list of objects in the sky that are so far away that their angle distances from each other do not change while the earth changes its position on the orbit. One therefore measures how much the nearer object changes its position relative to those very distant objects. Historically it was the first method to determine any star distances – and it was much more complicated to use than today.

3. Angular Size of Known Objects. If one can derive from solar and lunar eclipses that the diameter of the moon is one quarter of the (known) diameter of the earth, then one can determine the distance to the moon from its angular size of $\approx 0.5°$. Similarly, if one knows the absolute size of an object (in the sky) that is large enough to have an angular size then its distance is easily obtained. No single star is large enough. However, for some double stars one has been able to determine the diameter of their orbit around each other. The 1987 supernova illuminated a gas disk that is tilted towards us. The nearest edge showed this illumination first and from the observed(!) time difference between the illumination of the nearest edge and the farthest edge one easily obtained the size of this object. The most important example before 1987 was a star cluster named *Hyaden*. It is so close to us that photos taken at 10 year intervals show a movement of these stars against the distant sky. As in perspective drawings the tangents of these orbits converge to a point in the sky and the position of this convergence point tells us the direction of the movement in space. From the Doppler shift of the Fraunhofer lines one knows the radial velocity of this movement. Both informations give the absolute velocity and comparison with the angular velocity in the sky gives the absolute distance. – The data of the 1987 supernova are presently the best absolute distance measurements to get the luminosity comparison method started.

4. Red Shift. It was possible to determine the red shift of electromagnetic signals from objects of very different distance. The observations showed, that farther away objects have a larger red shift, in fact, the red shift grows almost linearly with the distance. (This is not true for nearby galaxies where rotations around each other are the dominant cause of frequency shifts.) The simplest explanation for such red shift is the Doppler effect. The observation of the red shift provided the first evidence of an expanding universe. – For many objects with large red shift it is not possible to determine their distance independently from the red shift. In those cases one takes the red shift itself as distance information, by extrapolating the expansion law.

*The descriptions of these distance measurements are obviously in pre-Einstein language. We need to redescribe them so that they fit into the world of Relativity Theory.*

Special Relativity Description of Distance Measurements

All electromagnetic information reaches an observer along its backwards light cone. Interpretation of such information will not give a distance immediately, instead, the most direct interpretation will place the observed star somewhere on the observer's backwards light cone. We can intersect that light cone with hyperplanes that are parallel to the simultaneity space of the observer. We can *compute* distances by multiplying the time distance of each simultaneity space with the speed of light and we *associate these distances to the star positions* on the backwards light cone. Next consider a different observer at the vertex of the above light cone. It has a different time unit vector and therefore its simultaneity spaces are tilted against the simultaneity spaces of the first observer. While we will see that the geometric positions on the backwards light cone remain the same the result of the tilt is that stars at equal distance for the first observer are not at equal distance for the second. This means that all distances in the universe change as the earth orbits the sun (although by *considerably less* than the error margin of astronomical distance measurements).

1. It is easy to adjust the luminosity comparison from a Newtonian description to a relativistic one: For a given observer the affine parametrizations of the light rays are given by the intersection of the light cone with the simultaneity spaces. Also, these intersections are spheres and their area increases quadratically with the (time) distance. Therefore, if we have placed one star on the backwards light cone and observe a second one of the same type, but one quarter as bright, then it is clear where on the light cone to place the second one. (Recall that the ratio of distances in *different* space directions depends on the observer at the light cone vertex.)

2. We represent the base line for the parallax measurement by a small segment in the simultaneity space of the observer – or, in view of the smallness of available base lines, by a tangent vector to the manifold orthogonal to the observer's time unit vector. Observation of the star means: connect the world line of the star and the points of the base line by null geodesics $c_\epsilon(s)$. Or again, the *infinitesimal version* is the Jacobi field $J(s) = \frac{d}{d\epsilon}c_\epsilon(s)$ along the null geodesic $c(s)$ from the star, $s = S$, to the observer $s = O$. $c(S)$ represents the star and $J(S)$ is tangential to the stars world line. $c(O)$ represents the observer and $J(O)$ is tangential to the base line in the simultaneity space of the observer, hence $g(c'(O), J(O)) = 0$. From (5.) above we obtain that $J(s)$ is actually tangential to the light cone, so that $J(S) = 0$. A parallax measurement determines how the direction to the star changes along the base line, i.e. one determines $\frac{D}{d\epsilon}\frac{d}{ds}c|_{\epsilon=0,s=O} = \frac{D}{ds}J(O)$. In Special Relativity Jacobi fields are linear ($J'' = 0$), therefore $J(O)$, $\frac{D}{ds}J(O)$ and the already fixed affine parametrization determine the value $s = S$ where $J(S) = 0$, i.e. the parallax measurement determines the point on the backwards light cone where we observe the star. (In retrospect, a change of the affine parametrization of the null geodesic changes the value of $S$, but not the position of the zero on the light cone.)
What is the result of a parallax measurement of a different observer $O'$ at $c(O)$? Let us extend the null geodesics from the star, $s \mapsto c_\epsilon(s)$, until they meet the rest space of $O'$ – in other words: to the extent that different observers can do this, $O$ and $O'$ use the same base line. Of course, the zero of the Jacobi field does not change, so tat $O'$, even

though measuring a different distance, places the star at the same point on the backwards light cone. The geometric picture is therefore better behaved than the communication of distances. (We could also have argued: in the quotient geometry along the light cone the two observers determine the same Jacobi field $[J](s)$.)

3. In the measurement of the angular size of known objects we use tangential Jacobi fields along the light cone (of the star or the observer) with $J(O=0)=0$ and $J'(O)$ being measured. We compute $s=S$ so that $J(S)$ has the size of the known object. Again. the point $c(S)$ on the light cone remains the same if we change the affine parametrization from $c(s)$ to $c_{new}(s) := c(\lambda s): \ c'_{new}(0) = \lambda c'(0), \ S_{new} = S/\lambda, \ c(S) = c_{new}(S_{new})$. – Note that the known object should be orthogonal to the direction (in space) of the connecting light ray, otherwise $J(S)$ is not tangential to the light cone.

4. I repeat the assumptions for using red shift as a distance measure: (i) The red shift is a Doppler frequency shift caused by relative velocity that increases linearly with distance. (ii) For objects for which only red shift measurements (and **no** independent distance measurements) can be obtained one still assumes the linear increase of velocity with distance. There is no problem in rephrasing these assumptions in terms of positions on the backwards light cone instead of in terms of distances.

## Distance Measurements in the Presence of Curvature

This situation is dealt with in two ways:

a) One has an explicit model for a planetary or cosmological situation. Then one can replace the linearity of Jacobi fields (from $J''=0$) by the known ODE
$$\frac{D}{ds}\left(\frac{D}{ds}[J](s)\right) + [R]([J](s), c'(s))c'(s) = 0$$
to compute the position of S on the light cone.

b) One has no information about the 4-dimensional geometry. Then one interprets the measurements as in the flat ($R=0$) case. But one can argue in which direction the result differs from the true position of S, if one makes assumptions about the sign of the eigenvalues of $[R]$.

We postpone a) until we have explicit models. As for b), if the eigenvalues of $[R]$ are positive the Jacobi field stays **below** the tangent at O. The parallax distances are therefore computed larger than they really are with the linear evaluation method. The distances which are linearly computed from the angle size of known objects are smaller than the real distances. – If the eigenvalues of $[R]$ are negative, then the Jacobi field stays **above** its tangent at O. Therefore the situation is reversed: linearly computed parallax distances are too large and linearly computed angle size distances are too small.
If the two eigenvalues of $[R]$ have different sign, then the parallax distance depends on the angle position (orthogonal to the direction to the star) of the base line and the angle size distance depends on the angle position of the known object – in particular, a circular disk is not observed as circular.

## Schwarzschild Geometry

From Planetary Systems to Black Holes

Newtonian celestial mechanics starts with the two-body problem where two masses move around their center of mass on Kepler orbits. No such solution is known in General Relativity. What is known, models a heavy central star and infinitesimal planets (they are so small that they do not influence the geometry). This model is called *the Schwarzschild solution*.

One cannot look for such a planetary model without mentioning the *Einstein equations*. These connect the geometric model with properties of the matter that exists in this model. One assumes that one knows the physical properties of the matter so well that one can write down, at each point in the rest system of the matter, a so called stress energy tensor $T$. In the present situation we assume vacuum outside the central star and vacuum is modelled by $T = 0$. (We cannot model the inside of the star where matter in complicated motion is present.) So we dealt with the physics side of the Einstein equations in a trivial way in this first example.

The other, the geometric side of the Einstein equations is independent of the matter. One needs the divergence free part of the Ricci tensor, the so called

Einstein tensor $\quad G := Ric - \frac{1}{2}\text{trace}\,(Ric)\text{id}\,.$

Note $\quad \text{trace}\,(G) = \text{trace}\,(Ric) - \frac{4}{2}\text{trace}\,(Ric) = -\text{trace}\,(Ric).$

In addition there is a parameter $\Lambda$, called *the Cosmological Constant*, in the Einstein equations. Einstein wrote the equations first with $\Lambda$, later without $\Lambda$. Cosmologists computed their main models first without $\Lambda$, presently with $\Lambda$. But the data fit of the astronomers to obtain $\Lambda$ from observations and the opinion of elementary particle physicists about the size of $\Lambda$ are more orders of magnitude apart than at any other disagreement in the history of physics. The famous Schwarzschild geometry has $\Lambda = 0$, but I will keep the parameter since two pioneer rockets from the seventies which have left the solar system in opposite directions deviate from their computed paths in an unexplained way. Finally, here are the

**Einstein equations:** $\quad 8\pi T = G + \Lambda \cdot \text{id}\,,$

Vacuum case: $\quad T = 0 \Rightarrow Ric = \Lambda \cdot \text{id}\,.$

Of course, the vacuum equations do not define anything specific. Further assumptions go into the Schwarzschild geometry: One wants a stationary and spherically symmetric geometry. Stationary means: time translation $t \to t + const$ is a Lorentz isometry. Spherically symmetric means: $SO(3)$ acts isometrically on 2-spheres obtained by setting radial coordinate and time coordinate to constants.

Ansatz: $\quad M := (a, b) \times \mathbb{S}^2 \times \mathbb{R}$

$$ds^2 := d\rho^2 + G^2(\rho)d\sigma^2 - F^2(\rho)dt^2,$$

where $d\sigma^2$ is the standard metric on $\mathbb{S}^2$.

Remark. The physics literature prefers a coordinate $r := G(\rho),\ dr = G'(\rho)d\rho$. Our *Ansatz* is closer to standard formulas in Differential Geometry, and the switch to the physics coordinates can trivially be made at the end.
Remark. The following determination of the Ricci tensor is done with the intention to explain as much as possible about this remarkable geometry, not as a computational exercise.

LEMMA. Let $N$ be the normal of a totally geodesic hypersurface, then $N$ is an eigenvector of the Ricci tensor.
Proof. For every tangent vector $X$ of the hypersurface we have $D_X N = 0$. Hence we have for every pair of tangent vectors $X, Y$ that $D^2_{X,Y}N = 0$ hence $D^2_{X,Y}N - D^2_{Y,X}N = R(X,Y)N = 0$. The curvature tensor symmetries imply $g(R(N,Y)Y, X)) = 0$, so that $R(N,Y)Y \in \mathbb{R}\cdot N$. Finally, the trace-definition of Ricci gives $Ric(N) = \lambda_N N$.

The easiest way to get totally geodesic hypersurfaces is as fixed point sets of isometries. Our Schwarzschild-Ansatz implies that time reflections are isometries and so are reflections in great circles of $\mathbb{S}^2$. We can therefore apply the previous Lemma to the timelike unit tangent vectors $e_4$ of the last factor, the $t$-factor, in $M := (a,b)\times\mathbb{S}^2\times\mathbb{R}$. Similarly we can apply the Lemma to all spacelike unit tangent vectors $e_\sigma$ of the factor $\mathbb{S}^2$. Finally, since the three factors are pairwise orthogonal, also the unit tangent vectors $e_1$ of $(a,b)$ must be eigenvectors of Ricci. Therefore we have proved:

*The tangent spaces of the three factors of $M$ are eigenspaces of the Ricci tensor of $M$.*

To obtain the eigenvalues we compute some curvature values from the Jacobi equation: $\frac{D}{ds}(\frac{D}{ds}J(s)) = -R(J,c')c'(s)$, for the remaining ones we use the Gauss equation and the following

LEMMA. If one has a family of parallel hypersurfaces (so that the vector field $N$ of unit normals has *geodesic* integral curves) then one can compute the shape operator $S$ of these hypersurfaces from Jacobi fields $J$ along the normal geodesics as follows:

$$S\cdot J = J'.$$

Proof. Let $c(t)$ be a curve in one of the parallel hypersurfaces and $\dot c(0) = X$. Let $s \mapsto c(s,t)$ be the family of normal geodesics. Observe $N(c(s,t)) = \frac{d}{ds}c(s,t)$ and $J_t(s) := \frac{d}{dt}c(s,t)$ are variations of geodesics, hence Jacobi fields along the normal geodesics. Of course $J_0(0) = X$. By definition of the shape operator we have

$$S\cdot J_t(s) = S\cdot\frac{d}{dt}c(s,t) := \frac{D}{dt}N(c(s,t)) = \frac{D}{dt}(\frac{d}{ds}c(s,t)) = \frac{D}{ds}(\frac{d}{dt}c(s,t)) = \frac{D}{ds}J_t(s).$$

The hypersurfaces $\rho = const$ are such a family of parallel hypersurfaces because the $\rho$-lines are geodesics (either from the form of the metric or by intersection of totally geodesic subspaces). These hypersurfaces have the product metric
$G^2(\rho)d\sigma^2 - F^2(\rho)dt^2$
with hypersurface curvature $1/G^2$ tangential to $\mathbb{S}^2$ and hypersurface curvature 0 for the $e_\sigma\wedge e_4$-planes. The eigenvalues of the shape operator are obtained with the lemma, where the needed Jacobi fields are restrictions of Killing fields. These Jacobi fields are obtained

from variations in 2-dimensional totally geodesic subspaces, which implies $J(s)/|J(s)|$ is a *parallel* field. Therefore we do not need Christoffel symbols to compute $(s = \rho)$

$$\frac{D}{ds}(\frac{D}{ds}J(s)) = |J(s)|'' \cdot \frac{J(s)}{|J(s)|} = -R(J(s), c'(s))c'(s), \quad c'(s) = e_1.$$
$$|J_4(\rho)| = F(\rho), \quad |J_\sigma(\rho)| = G(\rho).$$
$$R(e_4, e_1)e_1 = -\frac{F''(\rho)}{F(\rho)}e_4, \quad R(e_\sigma, e_1)e_1 = -\frac{G''(\rho)}{G(\rho)}e_\sigma.$$
$$S \cdot e_4 = \frac{F'}{F}e_4, \quad S \cdot e_\sigma = \frac{G'}{G}e_\sigma.$$
$$R(e_2, e_3)e_3 = (\frac{1}{G^2} - (\frac{G'}{G})^2)e_2, \quad R(e_3, e_2)e_2 = (\frac{1}{G^2} - (\frac{G'}{G})^2)e_3,$$
$$R(e_4, e_\sigma)e_\sigma = 0 - (\frac{G'}{G})(\frac{F'}{F})e_4, \quad R(e_\sigma, e_4)e_4 = 0 - (-\frac{F'}{F})(\frac{G'}{G})e_\sigma.$$

Note that in the Gauss equations, because of $g(Sy, y)Sx = \lambda_y g(y, y)\lambda_x x$, it matters which vectors are timelike and which are spacelike.
From the above curvature tensor data we obtain the eigenvalues of Ricci:

$$Ric(e_1) = \lambda_1 e_1 = (-\frac{F''}{F} - 2\frac{G''}{G}) \cdot e_1,$$
$$Ric(e_\sigma) = \lambda_\sigma e_\sigma = (\frac{1}{G^2} - (\frac{G'}{G})^2 - \frac{G''}{G} - \frac{F'G'}{FG}) \cdot e_\sigma,$$
$$Ric(e_4) = \lambda_4 e_4 = (-\frac{F''}{F} - 2\frac{F'G'}{FG}) \cdot e_4.$$

Einstein vacuum equations: $\Lambda = \lambda_1 = \lambda_\sigma = \lambda_4.$

The Schwarzschild geometry is obtained by solving this ODE-system.

$$(*1) : \lambda_1 = \lambda_4 \qquad \Leftrightarrow \frac{F'}{F} = \frac{G''}{G'} \Leftrightarrow (\frac{F}{G'})' = 0 \Leftrightarrow \frac{F}{G'} = const.$$

By scaling the t-coordinate we can have $const = 1$, hence

$$F = G'.$$

This leaves only one function to be determined. Setting all three eigenvalues equal and inserting $F = G'$ gives a third order ODE for $G$ which we can integrate twice for a first order ODE for $G$ that contains two parameters, one is the cosmological constant, the other will be called $m$ for mass of the central star.

$$(*2) : \lambda_\sigma = (\lambda_4 + \lambda_1)/2 \iff 0 = \frac{F''}{F} + \frac{1}{G^2} - (\frac{G'}{G})^2 = \frac{G'''}{G'} + \frac{1}{G^2} - (\frac{G'}{G})^2$$
$$\iff \left(\frac{-1}{G^2} + (\frac{G'}{G})^2 + 2\frac{G''}{G}\right)' = 2\frac{G'}{G}\left(\frac{G'''}{G'} + \frac{1}{G^2} - (\frac{G'}{G})^2\right) = 0.$$

If we compare the obtained first integral with $\lambda_\sigma$, we find that the value of this constant function is $-\Lambda$:

$$(\frac{-1}{G^2} + (\frac{G'}{G})^2 + 2\frac{G''}{G}) = -\Lambda.$$

We add the third order ODE and multiply by $G^2G'$ to get $\big(\frac{G'''}{G'} + 2\frac{G''}{G}\big)G^2G' = -\Lambda G^2G'$, hence another constant function:

$$\big(G^2G'' + \frac{\Lambda}{3}G^3\big)' = 0.$$

Define $$m := \big(G^2G'' + \frac{\Lambda}{3}G^3\big),$$

and observe $$m \ = \ G^2G'' + \frac{\Lambda}{3}G^3 = \frac{G}{2}(1 - G'^2) - \frac{\Lambda}{6}G^3.$$

So we arrived at the desired first order ODE for $G$:

$$\boxed{G'^2 = 1 - \frac{2m}{G} - \frac{\Lambda}{3}G^2}$$

Note that this ODE implies the third order ODE and hence all other used identities.

Finally let us make the change to the historic coordinates $r := G(\rho), dr = G'(\rho)d\rho$ and recall that the historic Schwarzschild solution has $\Lambda = 0$. The metric is

$$\boxed{ds^2 = (1 - \frac{2m}{r} - \frac{\Lambda}{3}r^2)^{-1}dr^2 + r^2d\sigma^2 - (1 - \frac{2m}{r} - \frac{\Lambda}{3}r^2)dt^2.}$$

We have computed above the Jacobi part of the curvature tensor in terms of $F, G$, now we use the ODE to compute these curvatures in terms of $m, \Lambda$. Note $(G'/G)(F'/F) = G''/G = m/G^3 - \Lambda/3$, $(1 - G'^2)/G^2 = 2m/G^3 + \Lambda/3$ and $-F''/F = -G'''/G' = (1 - G'^2)/G^2 = 2m/G^3 + \Lambda/3$. To obtain in addition to the six values already listed also $R(e_1, N)N$ with $N$ one of the totally geodesic hypersurface normals $e_4, e_\sigma$ used above, note that we proved already $R(e_1, N)\tilde{N} = 0$ if $N \perp \tilde{N}$ are any two of those normals. This says that $R(e_1, N)N$ is a multiple of $e_1$ and $g(R(e_1, N)N,\ e_1)$ is already known. This gives the following list

$$\begin{aligned}
&R(e_4, e_1)e_1 = (2m/G^3 + \Lambda/3)e_4, \quad R(e_\sigma, e_1)e_1 = (-m/G^3 + \Lambda/3)e_\sigma,\\
&R(e_2, e_3)e_3 = (2m/G^3 + \Lambda/3)e_2, \quad R(e_3, e_2)e_2 = (2m/G^3 + \Lambda/3)e_3,\\
&R(e_4, e_\sigma)e_\sigma = (-m/G^3 + \Lambda/3)e_4, \quad R(e_\sigma, e_4)e_4 = (+m/G^3 - \Lambda/3)e_\sigma,\\
&R(e_1, e_\sigma)e_\sigma = (-m/G^3 + \Lambda/3)e_1, \quad R(e_1, e_4)e_4 = -(2m/G^3 + \Lambda/3)e_1.
\end{aligned}$$

Check signs by computing $x \mapsto \sum_i R(x, e_i)e_i/g(e_i, e_i) = \Lambda x$. For $m = 0$ we get the curvature of the last example in lecture 2, $R(X, Y)Z\big|_{m=0} = (\Lambda/3)(g(Y, Z)X - g(X, Z)Y)$. Even the coordinates are almost the same: define a new coordinate $\alpha$ by $r =: \sqrt{\frac{3}{\Lambda}}\sin(\sqrt{\frac{\Lambda}{3}}\alpha)$ then $d\alpha^2 = dr^2/(1-\frac{\Lambda}{3}r^2)$ – which is that last example, with curvature $\Lambda/3$.

**Light Cone Curvatures**

First we show that the trace of the curvature $[R]$ of the quotient geometry on the tangent spaces of the light cones is zero.
Let $\vec{n}_1$ be tangent to the light ray in question. Extend this to a basis $\{\vec{e}_1,\vec{e}_2,\vec{n}_1,\vec{n}_2\}$ such that each $\vec{e}_j$ is a unit vector orthogonal to the other three and such that the $\vec{n}_j$ are null vectors with $g(\vec{n}_1,\vec{n}_2)=1$. In such a basis we have:

$$\vec{X} = x^1\vec{e}_1 + x^2\vec{e}_2 + y^1\vec{n}_1 + y^2\vec{n}_2 \quad\Rightarrow$$
$$x^1 = g(\vec{X},\vec{e}_1),\ x^2 = g(\vec{X},\vec{e}_2),\ y^1 = g(\vec{X},\vec{n}_2),\ y^2 = g(\vec{X},\vec{n}_1)$$

We compute the trace of an endomorphism $A$ by computing the above coefficients for $\vec{X} := A\vec{Y}$, where $\vec{Y}$ runs through the basis. Then

$$\mathrm{trace}\,A = g(A\vec{e}_1,\vec{e}_1) + g(A\vec{e}_2,\vec{e}_2) + g(A\vec{n}_1,\vec{n}_2) + g(A\vec{n}_2,\vec{n}_1) \quad \text{hence}$$
$$ricci(\vec{X},\vec{Y}) := \mathrm{trace}\,(\vec{Z}\mapsto R(\vec{Z},\vec{X})\vec{Y})$$
$$= g(R(\vec{e}_1,\vec{X})\vec{Y},\vec{e}_1) + g(R(\vec{e}_2,\vec{X})\vec{Y},\vec{e}_2) + g(R(\vec{n}_1,\vec{X})\vec{Y},\vec{n}_2) + g(R(\vec{n}_2,\vec{X})\vec{Y},\vec{n}_1).$$

Since $ricci(\vec{X},\vec{Y}) = \Lambda g(\vec{X},\vec{Y})$ we have $ricci(\vec{n}_1,\vec{n}_1)=0$ and therefore

$$0 = g(R(\vec{e}_1,\vec{n}_1)\vec{n}_1,\vec{e}_1) + g(R(\vec{e}_2,\vec{n}_1)\vec{n}_1,\vec{e}_2) + 0 + 0$$
$$= g([R]([\vec{e}_1],\vec{n}_1)\vec{n}_1,[\vec{e}_1]) + g([R]([\vec{e}_2],\vec{n}_1)\vec{n}_1,[\vec{e}_2]) = \mathrm{trace}\,[R].$$

In particular, the eigenvalues of $[R]$ agree up to sign.
The eigenvalues of $[R]$ will depend on $\vec{n}_1$. We put

$$\vec{n}_1 = (x^\rho, x^\sigma, 0, x^t) \quad \text{with} \quad (x^\rho)^2 + G(\rho)^2(x^\sigma)^2 - F(\rho)^2(x^t)^2 = 0.$$

Next we choose two tangent vectors to the light cone through $\vec{n}_1$

$$\vec{u} := (0,0,x^3,0) \perp \vec{n}_1,\ \vec{v}\perp\vec{u},\vec{n}_1$$

Recall that $R(\vec{X},\vec{Y})\vec{Z}=0$ if we insert three orthogonal vectors tangent to the factors of $M=(a,b)\times\mathbb{S}^2\times\mathbb{R}$. So we get

$$R(\vec{e}_3,\vec{n}_1)\vec{n}_1 = (x^\rho)^2 R(\vec{e}_3,\vec{e}_1)\vec{e}_1 + G(\rho)^2(x^\sigma)^2 R(\vec{e}_3,\vec{e}_2)\vec{e}_2 + F(\rho)^2(x^t)^2(\vec{e}_3,\vec{e}_4)\vec{e}_4$$
$$= \frac{3m}{G}(x^\sigma)^2\cdot\vec{e}_3$$

Since we had guessed an eigenvector ($\vec{e}_3$), this computation gives the eigenvalue. Note that the cosmological constant dropped out. The other eigenvalue is the negative of this one. This result says for example: If the light ray from the source has $x^\sigma \neq 0$ then we see circular objects at the source *not* circular. Or in other words, the distance measurement by angular size of a known object changes if the object is rotated by 90° (around the line of sight).

## Interesting Observers

The case of non-vanishing $\Lambda$ in the Schwarzschild geometry I have not seen discussed by astronomers. For the comparison with a Newtonian planetary system we will therefore usually assume $\Lambda = 0$. We will study this geometry by looking at it from the point of view of the following observers:

1. The Killing Observers. We set all the spacial coordinates constant and look at the world lines

$$\gamma(s) := (\rho_0, \sigma_0, F(\rho_0)^{-1} s) \in (a,b) \times \mathbb{S}^2 \times \mathbb{R}.$$

As curves these are the integral curves of the Killing field $X = (0,0,0,1)$ whose flow is coordinate-time-translation. We have reparametrized these curves with proper time: $g(\gamma'(s), \gamma'(s)) = -1$. In space such an observer would have to use a rocket to keep the spacial coordinates constant. If we model the space outside of a star by the Schwarzschild Geometry, then the observer could simply sit on the surface of the star.

2. Circling Observers. The world lines of these observers have the radial coordinate constant and the $\mathbb{S}^2$-coordinate runs on a great circle.

$$\gamma(s) = (\rho, \sigma_\rho(s), 0, \tau(\rho)\cdot s), \quad \gamma'(s) = (0, \omega(\rho), 0, \tau(\rho)) \quad \text{with}$$
$$-1 = g(\gamma'(s), \gamma'(s)) = G^2(\rho)\omega(\rho)^2 - F^2(\rho)\tau(\rho)^2.$$

If these world lines are geodesics then they model for example (infinitesimally small) planets circling the Schwarzschild Geometry. Otherwise they need again the help of rockets to stay on these orbits. An airplane flying with constant speed above a great circle is also such a circling observer.

We first discuss the Killing observers. Although they are not inertial observers – we will later compute their acceleration – they have some reason to consider themselves at rest relative to each other: a) Since each world line is the orbit of a 1-parameter family of isometries each of them sees the others **not** changing their angular sizes in time. b) If we consider a light signal from one such observer to another and back, then the time translation isometries carry the round trip world lines of these light signals to later round trip world lines. In other words, the round trip travel times of light signals between Killing observers don't change in time. c) If a periodic signal makes this round trip then the signal returns with the frequency of emission, i.e. no Doppler shift indicates relative motion. Since the observers are Killing observers we can compute the red shift without explicitly knowing the null geodesic $\gamma(s)$ between them, because time translation generates their world lines and the joining null geodesics from the initial situation and the Jacobi field needed for the red shift computation simply is the restriction of the Killing field $X = (0,0,0,1)$. The scalar product $g(X(\gamma(s)), \gamma'(s))$ is *constant* and the unit tangent vectors of source S and observer O are $u = F^{-1}(S)X(S),\ v = F^{-1}(O)X(O)$, therefore

$$1 + z = \frac{\omega_S}{\omega_O} = \frac{g(u,\gamma')}{g(v,\gamma')} = \frac{F(O)}{F(S)} = \frac{G'(O)}{G'(S)}.$$

The red shift on the return trip will cancel the red shift on the outgoing portion !!

Now we discuss the unexpected consequences of this computation. First, the source observer can tell the receiving observer *which frequency* the emitted signal has and then the receiving observer will see a red shift if his radial coordinate is larger and a blue shift otherwise! Therefore there is no way to interpret this frequency shift as a Doppler shift, because a Doppler shift is the same for source and receiver. Instead they have to come to the conclusion that between coordinate time slices $t = t_0$ and $t = t_1$ **more** time passes on the world line of that Killing observer that has the larger $G(\rho)$ than on the world line of the other. Below we compare the Schwarzschild Geometry with Newton's theory. This will show that we just gave the relativistic explanation of the Pound and Rebka experiment described in the first lecture.

This difference of the passing of time has unpleasant consequences for distance measurements: The round trip **coordinate** travel time of a light signal between Killing observers with different $\rho$ is the same in both directions. But since the ratio between coordinate time and proper time is **different** for these two observers they find that the light signal travel time distance is **not** the same for the two. – In the section on light cone curvatures we saw that also the distance by angle size of known objects depends on how one performs the measurement. So we must accept that distance measurements in the Schwarzschild geometry only give numbers which we agreed to call distances but which no longer behave as we expect distances to behave. – Killing observers with the same $\mathbb{S}^2$-coordinate can be said to be on the same radius from the star. The curve $c(s) = (s, \sigma_0, t_0)$ is a geodesic that is, moreover, orthogonal to all the world lines $\gamma_s(t) := (s, \sigma_0, t_0 + t)$ of the Killing observers on this radial line. Therefore these observers could justify to consider this radial line as in their (common) simultaneity space. However, Einstein's definition via light rays meeting in the middle does not work: two light signals emitted from two points on such a radial curve towards each other do **not** meet in the middle but closer to the interior observer.

Finally, observe that the redshift between Killing observers goes to infinity as the source approaches a point where $F(S) = 0$, at $G = 2m$ if $\Lambda = 0$. Signals that have undergone infinite red shift can no longer be received. This is the first reason why the deep interior of the Schwarzschild geometry is called a **Black Hole**.

### Acceleration of World Lines

A world line $\gamma(s)$ that is parametrized by proper time $(g(\gamma'(s), \gamma'(s)) = -1)$ has the acceleration $\frac{D}{ds}\gamma'(s)$ in the rest space of $\gamma'(s)$, and this is the acceleration which an observer on this world line experiences. We have to compute these accelerations.

On the underlying product manifold $M = (a, b) \times \mathbb{S}^2 \times \mathbb{R}$ we have the product metric $d\rho^2 + d\sigma^2 - dt^2$. We can work with its covariant derivative $D^\times$ without introducing local coordinates on $\mathbb{S}^2$. We denote by $\Gamma(.,.)$ the difference tensor between the Schwarzschild covariant derivative $D$ and $D^\times$:

$$D_X Y \;=\; D^\times_X Y + \Gamma(X, Y).$$

Again, $\Gamma$ is computed with the $(+,+,-)$ cyclic permutation trick:

$$\begin{aligned} T_Z(g(X,Y)) &= g(D_Z X, Y) + g(X, D_Z Y) \\ &= g(D_Z^\times X, Y) + g(X, D_Z^\times Y) + (D_Z^\times g)(X,Y). \end{aligned}$$

With the notation $X = (X^\rho, X^\sigma, X^t),\ g(X,Y) = X^\rho Y^\rho + G(\rho)^2 \langle X^\sigma, Y^\sigma \rangle - F(\rho)^2 X^t Y^t$ we have

$$\begin{aligned} &(D_Z^\times g)(X,Y) = 2G(T_Z G)\langle X^\sigma, Y^\sigma\rangle - 2F(T_Z F) X^t \cdot Y^t, \\ &T_Z G = G'(\rho)\cdot Z^\rho, \quad T_Z F = F'(\rho) \cdot Z^\rho \\ &g(\Gamma(X,Y), Z) = \frac{1}{2}\big( -(D_Z^\times g)(X,Y) + (D_X^\times g)(Y,Z) + (D_Y^\times g)(Z,X)\big) \end{aligned}$$

Finally, the difference tensor is

$$\boxed{\Gamma(X,Y) = \begin{pmatrix} FF'X^tY^t - GG'\langle X^\sigma, Y^\sigma\rangle \\ (G'/G)(X^\rho Y^\sigma + Y^\rho X^\sigma) \\ (F'/F)(X^\rho Y^t + Y^\rho X^t) \end{pmatrix}}$$

The first application is the radial acceleration of the Killing observers. Their world lines are

$$\gamma(s) := (\rho_0, \sigma_0, s/F(\rho_0)), \gamma'(s) = (0,0,1/F(\rho_0)),$$

$$\frac{D}{ds}(\gamma'(s)) = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} + \Gamma \begin{pmatrix} 0 & & 0 \\ 0 & , & 0 \\ 1/F(\rho_0) & & 1/F(\rho_0) \end{pmatrix} = \begin{pmatrix} (F'/F)(\rho_0) \\ 0 \\ 0 \end{pmatrix},$$

and more explicitly in the Schwarzschild geometry

$$\frac{F'}{F}(\rho) = \frac{G''}{G'}(\rho) = \frac{m/G^2 - (\Lambda/3)G}{\sqrt{1 - 2m/G - (\Lambda/3)G^2}}(\rho) = \Big|_{\Lambda=0} \frac{m}{G^2}\frac{1}{\sqrt{1-2m/G}}(\rho).$$

In the classical Schwarzschild case ($\Lambda = 0$) this says: For $G \gg 2m$, i.e. for very large radial coordinate function $G$, the acceleration needed to stay at a fixed place is $m/G^2$, resembling the Newtonian force of a central mass $m$. We also notice that distinctly before $G$ becomes zero, namely as $G$ approaches $2m$, the radial acceleration of the Killing observer becomes infinite. This will turn out to be the reason why the classical Schwarzschild geometry seems to terminate at $G = 2m$. Clearly the form of the metric is tuned to the Killing observer and it is no surprise that the classical expression for the metric does not extend beyond $G = 2m$. We will see below that also the circling observers support the interpretation of $m$ as the mass of the central object with $G$, at least for large values, behaving like a radial distance.

The cosmological constant $\Lambda$ is exceedingly small so that the deviation from the classical Schwarzschild geometry has not led to observed effects in our part of the planetary system. We can easily see from the formula, that for one very large value of $G$, for $G = \sqrt[3]{3m/\Lambda}$, the acceleration vanishes so that the Killing observer can stay in place without a rocket.

### Comparison of Radial Functions

In the $\Lambda = 0$ case we now compare the occurring radial functions. We show that, at least for large values, they do not differ much. One function is the arc length $\rho$ on the radial geodesics $\rho \mapsto (\rho, \sigma_0, t_0)$, recall, these are tangential to the infinitesimal rest spaces of the Killing observers. The other radial function $r = G(\rho)$ is preferred in the literature, its geometric distinction is that it gives the area of the $\mathbb{S}^2$-factor as: $area = 4\pi r^2$. When using $r$ the metric only makes sense for $r > 2m$, therefore we make the irrelevant normalization when integrating the ODE for $G$: $G(0) = 2m$.

Claim:

1.) $$r := G(\rho) \le R^{\#}(\rho) := \sqrt{4m^2 + \rho^2}$$

2.) $$G(\rho) \ge R_b(\rho) := \sqrt{4m^2 + \rho^2 \frac{0.5 + \rho}{1 + \rho}}$$

So indeed, saying that $r$ is large means the same as saying $\rho$ is large, and then $r \approx \rho$.

Proof.
For 1.) the idea is to prove $(R^{\#})'(\rho) \ge \sqrt{1 - 2m/R^{\#}(\rho)}$ and to note $G(0) = R^{\#}(0)$:

$$(R^{\#})'(\rho) = \frac{\sqrt{\rho^2}}{\sqrt{4m^2 + \rho^2}} = \sqrt{1 - \frac{2m}{R^{\#}(\rho)}} \cdot \sqrt{1 + \frac{2m}{R^{\#}(\rho)}} \ge \sqrt{1 - \frac{2m}{R^{\#}(\rho)}}.$$

Similarly for 2.) since the condition at the end is true:

$$R_b'(\rho) = \frac{1}{R_b(\rho)}\Big(\rho\frac{0.5 + \rho}{1 + \rho} + \rho^2 \frac{0.25}{(1 + \rho)^2}\Big) = \sqrt{1 - \frac{R_b^2 - \Big(\rho\frac{0.5+\rho}{1+\rho} + \rho^2\frac{0.25}{(1+\rho)^2}\Big)^2}{R_b^2}}$$
$$\le \sqrt{1 - \frac{2m}{R_b(\rho)}} \quad \Longleftarrow \quad (1 + \rho)(0.5 + \rho) \ge \Big(0.5 + \rho + \frac{0.25\rho}{1 + \rho}\Big)^2 \qquad Q.E.D.$$

### Circular Planetary Observers

For the world lines of circling observers we have (with $\sigma_\rho(.)$ a great circle in $\mathbb{S}^2$)

$$\gamma(s) = (\rho, \sigma_\rho(s), 0, \tau(\rho) \cdot s), \quad \gamma'(s) = (0, \omega(\rho), 0, \tau(\rho)) \quad \text{with}$$
$$-1 = g(\gamma'(s), \gamma'(s)) \; = \; G^2(\rho)\omega(\rho)^2 - F^2(\rho)\tau(\rho)^2.$$

The world line of an infinitesimal planet is in addition geodesic, i.e.

$$\frac{D}{ds}\gamma'(s) = \begin{pmatrix}0\\0\\0\end{pmatrix} + \Gamma(\gamma'(s),\gamma'(s)) = \begin{pmatrix}FF'(\rho)\tau(\rho)^2 - GG'(\rho)\omega(\rho)^2\\0\\0\end{pmatrix} \overset{(!)}{=} \begin{pmatrix}0\\0\\0\end{pmatrix}$$

The **geodesic condition**, using $G' = F$, therefore is:

$$GG''(\rho)\tau(\rho)^2 - G(\rho)^2\omega(\rho)^2 = 0, \quad G'(\rho)^2\tau(\rho)^2 - G(\rho)^2\omega(\rho)^2 = +1,$$

$$\tau(\rho)^2 = (G'(\rho)^2 - GG''(\rho))^{-1} = (1 - \frac{3m}{G})^{-1},$$

$$\omega(\rho)^2 = \tau(\rho)^2\frac{G''}{G}(\rho) = \frac{(m/G^3 - \Lambda/3)}{(1-3m/G)}\Bigg|_{\Lambda=0} = \frac{m}{G^3}(1-\frac{3m}{G})^{-1}.$$

We compare these results with Kepler's third law – presently under the agreement that a *Killing observer signals* when orbits are completed (see next lecture):

$$(\text{Proper Period Time, Planetary Clock})^2 = \Big(\frac{2\pi}{\omega(\rho)}\Big)^2 = \frac{4\pi^2}{m}G^3\Big(1-\frac{3m}{G}\Big),$$

$$(\text{Coordinate Period Time})^2 = \Big(\frac{2\pi\tau(\rho)}{\omega(\rho)}\Big)^2 = \frac{4\pi^2}{m}G^3.$$

**2m:** $Tunit = 3.3{\cdot}10^{-6} sec, 1year = 9.46{\cdot}10^{12} Tunits,\ G_{earth} = 1.5{\cdot}10^8 km \Longrightarrow 2m = 3\,km.$
Coordinate time is the proper time of the Killing observer at infinity. If one planetary observer looks at the rotation of other planets then he does not observe the proper time of the others. The observed periods – as signaled by Killing observers – are the coordinate time periods corrected by the factor between the observers proper time and the coordinate time along his world line. Keplers third law, $periods^2 = factor \cdot orbitradius^3$ holds in the Schwarzschild geometry with the function $r = G(\rho)$ as orbit radius. The $factor$ differs by $(1 - 3m/G_{Observer})$ from the Newtonian case since for the
**planetary observer** we have:

$$proper\ planetary\ time = \big(coordinate\ time\big)\cdot\sqrt{1-\frac{3m}{G}}.$$

In Special Relativity we compute the relative velocity $v$ between two observers $X, Y$ from their scalar product $(1-v^2)^{-1} = \langle\langle X, Y\rangle\rangle^2$. We compute the **orbit speed** of a planetary observer from its scalar product with the Killing observer:

$$g(\gamma'_{Killing}\ ,\ \gamma'_{Planet})^2 = F(\rho)^2\tau(\rho)^2 = \frac{1-2m/G-(\Lambda/3)G^2}{1-3m/G} = \frac{1}{1-v^2},$$

and, if $\Lambda = 0$, we have, asymptotic to the Newtonian case, $v^2 = (m/G)(1-2m/G)^{-1}$. For the Killing observer the length of the orbit is: relative velocity $\times$ period time $= 2\pi G$.
If $\Lambda > 0$ then this $v^2$ is smaller by $\approx -(\Lambda/3)G^2$.
Note that $v$ approaches the speed of light as the radius $G$ approaches $3m$.

We have found enough agreement with the Newtonian results to call the Schwarzschild geometry a relativistic planetary system and we now look more carefully for non-Newtonian, for relativistic effects.

## Schwarzschild Geometry II

Falling Particles, Bending of Light, Shapiro Delay, Perihelion Advance, Spinning Planet, Kruskal Extension, Kerr Comments.

The orbits of particles in a Newtonian planetary system are obtained as follows: Conservation of angular momentum gives that the orbits are planar. These are described in polar coordinates by two functions $r(t), \varphi(t)$. We can use conservation of angular momentum again to eliminate $\dot{\varphi}(t)$ from the kinetic energy. Finally, conservation of energy gives a first order ODE for $r(t)$ that can be integrated to give the Kepler orbits.
The same strategy works in the Schwarzschild geometry. Conserved quantities are obtained from Killing fields. We first use the rotational Killing field that is orthogonal to the initial velocity of the particle and see that that orbit remains orthogonal to this field. In other words, the orbit can be described as

$$\gamma(s) = (\rho(s), \phi(s), \pi/2, t(s)),$$

where $\phi(s)$ traces the equator great circle of $\mathbb{S}^2$ (its polar angle is $\vartheta = \pi/2$).

$$\gamma'(s) = (\rho'(s), \phi'(s), 0, t'(s)),$$

where for $s$ to be proper time we have

$$-1 = g(\gamma'(s),\gamma'(s)) = \rho'(s)^2 + G(\rho)^2\phi'(s)^2 - F(\rho)^2 t'(s)^2.$$

Next we use the time translation Killing field and the the rotational Killing field tangential to the $\mathbb{S}^2$-component of $\gamma'$:

$$g(\begin{pmatrix} 0\\0\\0\\1 \end{pmatrix}, \gamma'(s)) = -F(\rho(s))^2 t'(s) = const =: -T, \quad F(\rho(s))^2 t'(s)^2 = \frac{T^2}{F(\rho(s))^2}.$$

$$g(\begin{pmatrix} 0\\1\\0\\0 \end{pmatrix}, \gamma'(s)) = G(\rho(s))^2 \phi'(s) = const =: \Omega. \quad G(\rho(s))^2 \phi'(s)^2 = \frac{\Omega^2}{G(\rho(s))^2}.$$

The function $G$ could, for large values, be identified with the "distance" from the center. Therefore we can view the last identity as conservation of angular momentum. The previous one relates coordinate time and proper time; if $T = 1$ and $\rho$ is large so that $F^2 \approx 1$ hence $\rho'(s)^2 + G(\rho)^2\phi'(s)^2 \approx 0$ then we can say: *The particle is at rest at infinity and proper time equals coordinate time.* Finally, from $-1 = g(\gamma'(s), \gamma'(s))$ we get in the case $T = 1$:

$$\rho'(s)^2 + G(\rho(s))^2\phi'(s)^2 = -1 + \frac{T^2}{1 - 2m/G} \approx \frac{2m}{G(\rho)},$$

which can be viewed as the analogue of Newtonian conservation of energy. Eliminating $\phi'(s), t'(s)$ with these conservation laws gives as in the Newtonian case a first order ODE for $\rho(s)$:

$$\rho'(s)^2 = -1 - \frac{\Omega^2}{G(\rho(s))^2} + \frac{T^2}{1 - 2m/G(\rho(s))}.$$

After so much similarity we want to see some differences to the Newtonian case. First we list the invariants for the already computed circular orbits ($G = const$):

$$\Omega^2 := \frac{mG}{1 - 3m/G}, \qquad \frac{\Omega^2}{G^2} = \frac{m/G}{1 - 3m/G} \ ,$$
$$T^2 := \frac{(1 - 2m/G)^2}{(1 - 3m/G)}, \quad \frac{T^2}{F^2} = 1 + \frac{m/G}{1 - 3m/G} \ .$$

We see that the invariants get arbitrarily large as $G$ approaches $3m$ and these invariants do not exist in the range $2m < G \leq 3m$: there are no particle orbits in this part of the geometry. We have already computed the speed of the orbiting particle relative to the Killing observer at the end of the last lecture as $v^2 = (m/G)/(1 - 2m/G)^{-1}$ and as $G$ approaches $3m$ this relative velocity approaches 1, the velocity of light. This explains why particles cannot orbit on radii $G \leq 3m$.

We now consider the right side, of the ODE for $\rho(s)$, as function of $G$:

$$H(G) := -1 - \frac{\Omega^2}{G^2} + \frac{T^2}{1 - 2m/G}, \qquad \Omega^2, T^2 \text{ fixed.}$$

The particle orbit can only exist where $H(G) \geq 0$ and where $H(G) = 0$ we must have $\rho' = 0$, i.e. a point of minimal or maximal radial coordinate. We see: if $T^2 < 1$ then the particle cannot escape to infinity, we have along its orbit $G \leq 2m(1 - T^2)^{-1}$. And if $T^2 > 1$ and $\rho'(s_0) > 0$ then $\rho(s)$ ($s \geq s_0$) will grow to infinity with $\rho'(\infty)^2 = T^2 - 1$. The velocity relative to the Killing observer at infinity is $v^2 = 1 - 1/T^2$.
On most circular orbits we have $T^2 < 1$. But in the range $4m > G > 3m$ we have $1 < T^2 < \infty$. We will show that from any point outside $G = 4m$ we can give particles initial data such that they fall asymptotically to one of the circles at radius $G_\infty \in (3m, 4m]$ – and of course vice versa: arbitrarily small perturbations of such a circular orbit can make the particle fly to infinity, in fact with velocities relative to the Killing observer at infinity that are arbitrarily close to the speed of light – in sharp contrast to the Newtonian situation. How is this done? Let the initial point have radial coordinate $G_0$ and let us aim for the circular orbit at $G_\infty \in (3m, 4m]$. We have to choose orbit invariants that are quite different from the invariants $\Omega_0$, $T_0$ of a circling particle at the initial value $G_0$

$$\Omega_\infty^2 := \frac{mG_\infty}{1 - 3m/G_\infty}, \quad T_\infty^2 := \frac{(1 - 2m/G_\infty)^2}{(1 - 3m/G_\infty)} \geq 1 > T_0^2.$$

Recall

$$0 = -1 - \frac{\Omega_\infty^2}{G_\infty^2} + \frac{T_\infty^2}{F_\infty^2}, \quad \text{i.e.: } H(G_\infty) = 0.$$

In fact, $H(G) = -1 - \Omega_\infty^2/G^2 + T_\infty^2(1 - 2m/G)^{-1}$ has a double zero at $G = G_\infty$ since
$$H'(G) = 2\Omega_\infty^2/G^3 - T_\infty^2(1 - 2m/G)^{-2}2m/G^2$$
is also zero at $G = G_\infty$. Finally, $H(G)(1 - 2m/G)G^3$ is a cubic polynomial with highest

coefficient $T_\infty^2 - 1$ and absolut term $2m\Omega_\infty^2$. Therefore H(G) has no further zero in $(3m, \infty)$ and is positive at infinity. Therefore we can indeed prescribe at the radial distance $G_0$ the orbital invariants of the circular orbit at $G_\infty$. In the outward direction the orbit leaves the system, in the inward direction the orbit cannot reach $G = G_\infty$, because then it would have to agree with the circular orbit. But $\rho'$ cannot change sign before $G = G_\infty$, so that the orbit (that started at $G_0$) has to spiral towards the circular orbit. – Indeed, we will later find exponentially growing. resp. decaying, Jacobi fields corresponding to the described situation.

The discussion of these exotic orbits is intended to show features that do not exist in a Newtonian system. If we choose a larger angular momentum invariant than $\Omega_\infty$ then the orbit will reach the return point $\rho' = 0$ at a radial value $G > G_\infty$ and the orbit will come out again as in the Newtonian case. If we choose the angular momentum invariant less than $\Omega_\infty$ then $\rho' = 0$ can **not** occur for $G \geq G_\infty$, but because of the term $T_\infty^2(1 - 2m/G)^{-1}$ a return point $\rho' = 0$ may never be reached and the orbit continues into the black hole. By contrast, in the Newtonian case only strictly radial orbits, $\Omega = 0$, do not return out (except if the star itself is in the way).

### What does “once around” mean?

Simultaneity causes a problem that is not immediately apparent because of our Newtonian training. If a particle circles the star and a Killing observer with the same value of the radial function $G$ observes the particle there is no problem: The world line of the circling particle crosses the world line of the waiting Killing observer periodically and the Killing observer says, *the orbiter completes one revolution between neighboring intersection points and the time length of my Killing world line between these intersections is the observed period duration.* And the Killing observer computes the length of the periodic orbit as $2\pi G$ from the relative velocity and the period duration. The orbiting observer could also say: *I have completed one orbit when I meet the Killing observer the next time.* However, from the orbiters point of view the Killing observer races towards him, and certainly, when two people run in opposite directions around a stadium, they will not say that they completed one circuit when they meet again. Moreover, when we imagine the orbit full of circling particles (like the rings of Saturn) then we should consider all their world lines. It seems also reasonable to agree that *the circling particles have completed one revolution if the world lines of* **all of them** *have been met by the Killing observer moving towards them.* (Certainly, if some thing flies past my window, I clock the time until it has met all the windows world lines.) And, the circling particles measure the distance to close neighbors in the rest space orthogonal to their world lines. What does the circling observer see?

Recall that we have for the relative velocity $v$ between circling observer and Killing observer $v^2 = m/G(1-2m/G)^{-1}$, $1-v^2 = (1-3m/G)(1-2m/G)^{-1}$. The world lines of the rotating particles fill a cylinder. One orthogonal trajectory of these world lines is the best approximation to a rest space because nearby particles are relative at rest. The length of the line at rest from one Killing world line once around the cylinder to the same Killing world line has length $2\pi G\sqrt{1-v^2}$ and the time length, squared, $T_K^2$, of that Killing world line segment is $T_K^2 = 4\pi^2(G^3/m)(1-2m/G)$. At this length the rest line for the rotating particles has not met all world lines of these particles. We extend it until it meets the world line through its initial point again, this larger length is: $2\pi G/\sqrt{1-v^2}$. The time length of this segment on an rotating particles world line is $T_R = T_K/\sqrt{1-v^2}$. Note that we also get for the rotating observer: orbit length = relative velocity × period time =

$$= v \cdot T_R = 2\pi G/\sqrt{1-v^2} = 2\pi G(1-2m/G)^{1/2}(1-3m/G)^{-1/2}.$$

This shows that the number of elementary particles that fit next to each other on such a rotating orbit goes to infinity as $G$ approaches $3m$. Note that the numbers $v$ and $T_R$ do not make sense for the rotating particles by themselves, they need a second observer. On the other hand, the number $2\pi G(1-2m/G)^{1/2}(1-3m/G)^{-1/2}$ only depends on the world line of a rotating particle, it is the geometric length in the Schwarzschild geometry of the curve on the cylinder that is tangential to all the infinitesimal rest spaces of the rotating particles.
To come back to the beginning of this discussion, why shouldn't the circling particles agree that they have completed one revolution when they see the stars at infinity in the same position? These stars at infinity are assumed at rest relative to the Killing observer at infinity. Each particle separately can take this definition of a complete revolution. But then there still is the notion of being at rest relative to infinitesimal neighbors, this notion defines the distance between neighboring world lines and therefore determines, how many elementary particles fit onto one circle of Saturn's rings – and when we want to discuss the strangeness of indefinite geometries then this number is more important than the undoubted convenience of observing the stars at infinity.

**Behaviour of Light**

The cosmological constant is too small to play a role nearer to the center. Since we understand the asymptotic Schwarzschild geometry far away from the star in principle from lecture 2 I would have preferred to leave $\Lambda$ in the following discussion. However, the coordinates which come from the symmetry assumptions of the classical Schwarzschild Ansatz are not very suitable to study the limit behavior of null geodesics. When $\Lambda = 0$ then the Schwarzschild geometry at infinity is Special Relativity with a preferred inertial observer (the Killing observer). Therefore we can place the stars at the sky of this preferred observer and discuss the deviation of light. If $\Lambda \neq 0$ I have not succeeded in describing the stars in the sky. Therefore the following assumes the classical Schwarzschild geometry, i.e. $\Lambda = 0$.

The world lines $\gamma$ of light signals are null geodesics, and we use the same Killing fields as for particles to get conserved quantities $\Omega, T$.

$$g(\gamma'(s),\gamma'(s)) = 0, \quad {\rho'}^2 + G(\rho)^2{\varphi'}^2 = F(\rho)^2{t'}^2, \quad \frac{D}{ds}\gamma'(s) = 0,$$

$$G(\rho(s))^2 \cdot \varphi'(s) = \Omega, \quad G(\rho(s))^2\varphi'(s)^2 = \frac{\Omega^2}{G(\rho(s))^2},$$

$$F(\rho(s))^2 \cdot t'(s) = T, \quad F(\rho(s))^2 t'(s)^2 = \frac{T^2}{F(\rho(s))^2}.$$

Null geodesics have no preferred affine parameter (like arc length) we may assume $T = 1$.

The geodesic equation is again reduced to a first order ODE for $\rho(s)$:

$$\rho'(s)^2 = \frac{1}{F(\rho(s))^2} - \frac{\Omega^2}{G(\rho(s))^2}.$$

**The Photon Sphere.** We will find circular orbits of light. As in the Newtonian case we do not get the velocity on a circular orbit from the constant of the motion. The covariant derivative is the same as for particles :

$$\frac{D}{ds}\gamma'(s) = (0,0,0) + \Gamma(\gamma'(s),\gamma'(s)) = (0,0,0),$$

$$\text{or:}\quad 0 = FF'(\rho)t'^2 - GG'(\rho)\varphi'^2 = \frac{F'}{F^3} - \frac{G'\Omega^2}{G^3}.$$

Now use $\rho' = 0$ hence $1/F^2 = \Omega^2/G^2$ and recall $F = G',\ G'' = m/G^2$ to get

$$\frac{m}{G} = 1 - \frac{2m}{G}, \quad \text{finally:} \quad \boxed{G = 3m,\ \Omega^2 = 27m^2.}$$

So indeed, at $G = 3m$ photons can circle the star!

**Black Hole again.** Light rays towards the star ($\rho'(s_0) < 0$) for which $\Omega^2$ is so small that $\rho'$ cannot become zero will fall into the star. Consider

$$\rho'^2(1-\frac{2m}{G})G^3 = G^3 - \Omega^2(G-2m) = (G-3m)^2(G+6m) - (\Omega^2 - 27m^2)(G-2m).$$

If $\Omega^2 = 27m^2$ then the incoming ray will be asymptotic to the photon sphere.
If $\Omega^2 > 27m^2$ then $\rho' = 0$ occurs at some value $G_{min} > 3m$, $\rho'$ changes sign and the light ray leaves the star. — Note $\Omega^2 = G_{min}^2/(1-2m/G_{min})$.
If $\Omega^2 < 27m^2$ then $\rho' \le -\epsilon$ and the ray disappears at $G = 2m$ at a finite value of its affine parameter $s$.

A similar discussion applies for light rays that start from inside the photon sphere in the outward direction, $G_0 \in (2m, 3m], \rho' > 0$.
If $\Omega^2 = 27m^2$ then the ray approaches the photon sphere asymptotically from inside.
If $\Omega^2 > 27m^2$ then $\rho' = 0$ occurs at some $G_{max} \in (G_0, 3m)$, the ray returns and falls into the star.
If $\Omega^2 < 27m^2$ then $\rho' > 0$ forever and this ray can leave the field of the star.
We compute, in the rest space of the Killing observer, the angle $\alpha$ between the radial direction and the direction of the asymptotic ray. For the asymptotic ray we have

$$G^2\varphi'^2 = \frac{27m^2}{G^2} \text{ and } \rho'^2 = \frac{1}{F^2} - G^2\varphi'^2,$$

hence:

$$\tan^2\alpha := \frac{G^2\varphi'^2}{\rho'^2} = \frac{1-2m/G}{G^2/(27m^2) - (1-2m/G)} \underset{G\to 2m}{\longrightarrow} = 0.$$

This says: The cone angle around the radial direction decreases from 90° to 0° as the radial coordinate of the initial point decreases from $G = 3m$ to $G = 2m$.

**Bending of Light.** The Schwarzschild geometry is asymptotically Special Relativity with the Killing observer at infinity as a distinguished inertial observer. We fix the angular coordinate so that $\varphi(s) = 0$ at the point of smallest radial coordinate $G = G_{min}$ along the ray. The differential equation shows that $\lim_{s\to\infty} \rho'(s) = 1$ and $\lim_{s\to\infty} G(\rho)\varphi'(s) = 0$ so that the ray leaves the star asymptotically in a fixed direction $\varphi_\infty$. Since the direction to the nearest point is orthogonal to the ray we have

**Deflection Angle**: $A = 2(\varphi_\infty - \pi/2) = 2\int_{G_{min}}^{\infty} \frac{d\varphi}{dG} dG.$

We insert $\quad \frac{dG}{d\rho} = \sqrt{1 - \frac{2m}{G}}, \quad \frac{d\rho}{ds} = \sqrt{\frac{1}{1-2m/G} - \frac{\Omega^2}{G^2}}, \quad \frac{d\varphi}{ds} = \frac{\Omega}{G^2}$

and abbreviate $\quad \epsilon := \frac{2m}{G_{min}}, \quad y := \frac{G_{min}}{G}, \quad dy = \frac{-G_{min}}{G^2} dG,$

$$\Omega^2 := \frac{G_{min}^2}{1 - 2m/G_{min}} = \frac{G_{min}^2}{1-\epsilon}$$

to obtain $\quad \frac{1}{2}A = \int d\varphi - \frac{\pi}{2} = \int_0^1 \frac{dy}{\sqrt{1-\epsilon-y^2+\epsilon y^3}} - \int_0^1 \frac{dy}{\sqrt{1-y^2}}.$

Next use $a := \sqrt{1-y^2},\ b := \sqrt{1-\epsilon-y^2+\epsilon y^3},\ \frac{1}{b} - \frac{1}{a} = \frac{a^2-b^2}{ab(a+b)},\ \frac{1-y^3}{1-y^2} = \frac{1+y+y^2}{1+y}$

to get $\quad \frac{1}{2}A = \ \epsilon \int_0^1 \frac{\big(1 + y^2/(1+y)\big)\ dy}{\sqrt{1-\epsilon-y^2+\epsilon y^3}\Big(1 + \sqrt{1-\epsilon-y^2+\epsilon y^3}/\sqrt{1-y^2}\Big)}$

The integral evaluates at $\epsilon = 0$ to 1 — giving $\frac{d}{d\epsilon}(\frac{1}{2}A)|_{\epsilon=0} = 1$ — since by partial integration

$$\int_0^1 \frac{y}{\sqrt{1-y^2}} \cdot (1 - \frac{1}{1+y})dy = 0 + \int_0^1 \frac{\sqrt{1-y^2}}{(1+y)^2} dy = -2\frac{\sqrt{1-y}}{\sqrt{1+y}}\Big|_0^1 - \int_0^1 \frac{dy}{\sqrt{1-y^2}} = 2 - \frac{\pi}{2}.$$

With this same integration we can actually get upper and lower bounds for the deflection angle (with the same $\frac{d}{d\epsilon}$ at 0) from a trivial estimate in the denominator (use $0 \le y \le 1$ and $-\frac{1}{2} + y^2(\frac{3}{2} - y) \le 0$ for $y \in [0,1]$ ):

$$\sqrt{(1-y^2)(1-\frac{3}{2}\epsilon)} \le \sqrt{1-\epsilon-y^2+\epsilon y^3} \le \sqrt{(1-y^2)(1-\epsilon)} \Longrightarrow$$
$$\frac{4\epsilon}{\sqrt{(1-\epsilon)}\cdot(1+\sqrt{1-\epsilon})} \le A \le \frac{4\epsilon}{\sqrt{(1-\frac{3}{2}\epsilon)}\cdot(1+\sqrt{1-\frac{3}{2}\epsilon})}.$$

This deflection of light can be measured with great accuracy for light or radio signals that pass near the sun. The sun diameter is $1.39{\cdot}10^6\ km$ hence $\epsilon = 4.3{\cdot}10^{-6}$ and $A = 2\epsilon = 1.77''$. The bending of light has become very important in cosmology because one can observe light from farther away sources such that the rays passed on the way to us very near some other galaxy and were bent by this galaxy's mass. The study of this *gravitational lensing* has led to conclusions about the deflecting masses: there is more gravitating mass then can be accounted for by visible matter like stars or dust.

## Shapiro Delay

Consider the observation of a pulsar from the orbiting earth with the pulsar lying in the plane of the earth orbit. When the radius vector from the sun to the earth is orthogonal to the direction of the pulsar, the earth is moving directly towards or away from the pulsar, resulting in maximal (blue and red) Doppler shift of all observed frequencies. When the direction to the pulsar passes near the sun, then the orbit velocity is almost orthogonal to the incoming radiation and Doppler shift is minimized. Of course the bending of the incoming ray around the sun is maximal in this position. We also observe the *Shapiro Delay*, a quite surprising phenomenon for Newtonian intuition. Looking back at the Schwarzschild formulas in an $\rho$-$t$-picture we recall that the light cones get steeper as $\rho$ decreases. This means that the null geodesics bridge more coordinate time when they travel some "distance" close to the star than farther out. Or in other words, the light ray seems to spend more time when it passes close to the central region. The effect is important because it is so unexpected in a Newtonian picture and so easily predicted from the Schwarzschild geometry. It can also be observed easily with accuracy. We want to apply the previous discussions to compute its size, but this requires to formulate the problem more precisely. First, the Schwarzschild traveling time of the light ray until it first meets the earth' orbit differs already from the Newtonian computation. Since I do not see how that difference could be observed I will ignore it and consider only the traveling light signal across the orbit of the earth. The relativistic prediction is obtained from the ODE of null geodesics, it gives how much coordinate time passes while the light ray crosses the earth orbit. (The proper time on the orbiting earth passes slower by a constant factor $(1-3m/G)$ which is irrelevant to explain the Shapiro delay.) For the Newtonian prediction we separate space and time, we compute the arc length of the projection of the null geodesic into the "space" orthogonal to the Killing observers (such space-slices $t = const$ are most commonly taken as the "spatial" geometry of the 4-dim Schwarzschild geometry). Division of this arc length by the velocity of light (=1) gives the Newtonian travel time. We use the same notation for light rays $\gamma(s)$, $\gamma' = (\rho', \varphi', t')$ as above. The Killing field is $(0,0,1)$ and the component of $\gamma'$ orthogonal to it is ${\gamma'}^{\perp} = (\rho', \varphi', 0)$.

$$g({\gamma'}^{\perp}, {\gamma'}^{\perp}) = (\rho')^2 + G(\rho)^2(\varphi')^2 = \frac{1}{F^2}$$

Newtonian travel time: $\displaystyle\int \frac{ds}{F}$

Relativistic travel time: $\displaystyle\int dt = \int t'(s)ds = \int \frac{ds}{F^2}$

As in the deflection of light computation we can insert

$$ds = \frac{ds}{d\rho}\frac{d\rho}{dG}dG, \quad \frac{d\rho}{ds} = \sqrt{\frac{1}{F^2} - \frac{\Omega^2}{G^2}}, \quad \frac{dG}{d\rho} = \sqrt{1 - \frac{2m}{G}} = F$$

to obtain more explicit expressions. However, the extra factor $F$ in the denominator of the relativistic travel time shows clearly how the Shapiro delay comes about.

Note that the above computation explains the delay without explicitly computing where on the orbit the rays are received. I find it instructive to also compute the red shift from the pulsar to the earth. As with the Shapiro delay the red shift is obtained without mentioning the corresponding points of reception on the earth orbit. Therefore we do not see that the red shift is a differentiated version of the delay. Actual observation of the pulsar peeks shows the frequency shift as the varying time distance between the peeks, but it also shows the integrated delay, because the pulsar signals are emitted at known equidistant time intervals.

We assume that the pulsar is at rest at infinity. Recall

$$F^2(\rho)t'(s) = 1, \quad G^2(\rho)\varphi'(s) = \Omega = \pm G_{min}(1 - 2m/G_{min})^{-1/2}, \quad \rho'(s)^2 = \frac{1}{F^2} - \frac{\Omega^2}{G^2}.$$

In particular, each $\Omega$ contains the information, how close to the sun that light ray passes. For the tangent vector $c'(s)$ of the circling earth we know:

$$c' = (0, \omega(\rho), 0, \tau(\rho)), \quad \omega^2 = \frac{m}{G^3}\frac{1}{1 - 3m/G}, \quad \tau^2 = \frac{1}{1 - 3m/G}$$

The red shift from the pulsar to the earth is given by Jacobi fields $J(s)$ along the light rays. As value of each Jacobi field at the source we can take the unit tangent vector at the world line of the source, namely $J(\infty) = (0, 0, 0, 1)$. Since this is the value of a Killing field we have its restriction to the light ray as a Jacobi field $K(s)$ along the light ray with the correct value at the source. However, we need a Jacobi field whose value at the earth is a multiple $\mu c'$ of the observers time unit vector. Such a Jacobi field $J(s)$ exists since we assume observation of the source from the earth. Both $K$ and $J$ come from variations of null geodesics, therefore both can be written as the sum of a *parallel* field and a field *tangential* to the light cone along $c$. Since both fields agree at the source, their parallel components agree, so that their difference must be tangential. This determines the factor $\mu$:

$$g\left(\mu \cdot (0, \omega, 0, \tau) - (0, 0, 0, 1),\ (\rho'(s), \varphi'(s), 0, t'(s))\right) = 0$$
$$\mu \cdot (G^2\omega\varphi' - F^2\tau t') = -F^2 t' = -1, \qquad \mu = \frac{1}{\tau - \omega\Omega}.$$

The frequency shift $frequency(source)/frequency(observer) = 1 + z$ therefore is

$$1 + z = \mu = \frac{1}{\tau - \omega\Omega} = \frac{\sqrt{1 - 3m/G}}{1 \mp \sqrt{(mG_{min}^2/G^3)/(1 - 2m/G_{min})}}.$$

As a check, the product of the two values for $G_{min} = G = G_{Earth}$ should give the square of the (blue) shift from infinity to a Killing observer on the earth orbit. Indeed

$$(\tau - \omega\Omega)(\tau + \omega\Omega)\Big|_{G_{min}=G} = (1 - \frac{2m}{G})^{-1}.$$

If one uses the above formulas not for the earth but for a planet circling a black hole then one should observe that the rays get bent more and more times around the center as $G_{min}$ approaches $3m$.

## Tidal Forces of Gravity

What can be observed when two particles travel with almost the same initial conditions next to each other? Both have timelike geodesic worldlines $\gamma(s), \gamma_\epsilon(s)$. We can imagine a family of geodesic world lines between them, differentiate with respect to $\epsilon$ and describe the second particle by a Jacobi field $J$ along $\gamma$. It is called separation vector field and satisfies the famous

**Jacobi equation:** $\frac{D}{ds}\big(\frac{D}{ds}J(s)\big) + R_{|\gamma(s)}\big(J(s), \gamma'(s)\big)\gamma'(s) = 0,$

or shorter: $J'' + R(J, \gamma')\gamma' = 0.$

Observe that we assumed geodesic world lines, which means: the neighboring particles do not feel any acceleration, no forces acting on them. However, when each one observes its neighbor they see that their separation vector shows **relative acceleration**. This relative acceleration, as the saying goes, is caused by *the tidal forces of gravity.* They should not be taken lightly: if such neighbors join each other with a stick then these tidal forces become very real. The closest Galilean moon of Jupiter is deformed so much by these forces that, astronomers believe, its volcanic activity is caused by this deformation heating. Our own moon is cold now, but its inside is not made of solid rock. When the molten material tried to solidify it was tidal forced into chunks of rock.

The above remarks are general. What can the Jacobi equation do for our understanding of the Schwarzschild geometry? One has to find a situation where the study of nearby geodesic world lines leads to observable predictions, preferably different from Newtonian predictions. When a single Newtonian planet circles the sun its orbit is a Kepler ellipse. In particular, the closest point to the sun, the *perihelion*, is at the same spot in space on every revolution. This is no longer the case if other planets, like Jupiter, perturb the situation. In case of Mercury the perihelion advances. But a careful analysis of all classical contributions explained only 532" per century of the observed 574" per century of Mercury's perihelion advance.

How are Jacobi fields related to this situation? A planet with an almost circular orbit can be described by a Jacobi field along a circular orbit. A relativistic contribution to the **perihelion advance** would be the existence of a periodic Jacobi field with *larger* period than the rotation period of the circular planet. This is what we will find.

Earlier we wrote the tangent vector of a circling geodesic world line as

$$\gamma'(s) = (0, \omega(\rho), 0, \tau(\rho)) \quad \text{with}$$
$$-1 = g(\gamma'(s), \gamma'(s)) \;=\; G^2(\rho)\omega(\rho)^2 - F^2(\rho)\tau(\rho)^2 \quad \text{and}$$
$$\tau(\rho)^2 = (1 - \frac{3m}{G})^{-1}, \;\; \omega(\rho)^2 = \frac{m}{G^3}(1 - \frac{3m}{G})^{-1}.$$

The simplest orthonormal basis, namely
$e_1 = (1, 0, 0, 0), e_2 = (0, 1/G, 0, 0), e_3 = (0, 0, 1/G, 0),\; e_4 = (0, 0, 0, 1/F)$ is adapted to the

Killing observer. We can use $e_1, e_3$ along $\gamma$, but the other two need to be adapted to the world line $\gamma$

$$\gamma' = (\frac{m}{G})^{1/2}(1-\frac{3m}{G})^{-1/2}\cdot e_2 + (1-\frac{2m}{G})^{1/2}(1-\frac{3m}{G})^{-1/2}\cdot e_4$$
$$f_2 := (1-\frac{2m}{G})^{1/2}(1-\frac{3m}{G})^{-1/2}\cdot e_2 + (\frac{m}{G})^{1/2}(1-\frac{3m}{G})^{-1/2}\cdot e_4$$
$$= (0, \frac{\tau F}{G}, 0, \frac{\omega G}{F})(\rho), \quad g(f_2, f_2) = F^2\tau^2 - G^2\omega^2 = +1.$$

We also recall the Christoffel symbols ($X = (x^\rho, x^\sigma, x^t),\ Y = (y^\rho, y^\sigma, y^t)$):

$$\Gamma(X,Y) = \begin{pmatrix} FF'X^tY^t - GG'\langle X^\sigma, Y^\sigma\rangle \\ (G'/G)(X^\rho Y^\sigma + Y^\rho X^\sigma) \\ (F'/F)(X^\rho Y^t + Y^\rho X^t) \end{pmatrix}.$$

As expected we have $\frac{D}{ds}e_3 = \Gamma(\gamma', e_3) = 0$, We will later see that such a parallel vector along a geodesic world line describes the axis of a rotating solid body with all its moments of inertia equal to each other.
We find

$$\frac{D}{ds}e_1 = \Gamma(\gamma', e_1) = (0, \frac{\omega G'}{G}, 0, \frac{\tau F'}{F}) = \sqrt{\frac{m}{G^3}}f_2 =: \omega_\varphi f_2.$$

Since we differentiate an orthonormal basis we also have

$$\frac{D}{ds}f_2 = -\sqrt{\frac{m}{G^3}}e_1 = -\omega_\varphi e_1.$$

The vector field $p(s) := e_1(s)\cos(\omega_\varphi s) - f_2(s)\sin(\omega_\varphi s)$ is a *parallel* field. This is similar to the Newtonian case where this field returns to its (radial) initial value after one complete rotation. Here the period is not quite right: It is plausible to let the Killing observer decide when the planet has completed one revolution (because we watch the perihelion advance from outside). We have computed how much planetary proper time (measured by $s$) passes until the Killing observer signals completion: $(2\pi/\omega_\varphi)(1-3m/G)^{1/2}$. This time is too short and $p$ is not yet radial. If the planetary observer waits until his rest space method says, the orbit is complete then too much time passed: $(2\pi/\omega_\varphi)(1-3m/G)^{-1/2}$. The different definitions of “completed orbit” lead to different experiments, each with a clear prediction from Relativity Theory.

We will use repeatedly that $R(e_i, e_j)e_k = 0$ when $i, j, k$ are pairwise different. From our earlier list of curvature values we now find

$$R(e_3, \gamma')\gamma' = \frac{m}{G^3}(1-\frac{3m}{G})^{-1}e_3 =: K_3e_3 = \omega^2 e_3,$$

where $\qquad \omega = \omega(\rho) \quad$ as in $\gamma' = (0, \omega, 0, \tau)$,

which gives the Jacobi fields

$$J_3(s) = (A\cos(\omega s) + B\sin(\omega s))\cdot e_3(s).$$

They describe tilted neighboring circular orbits. In this case the period of the Jacobi field agrees with what the Killing observer calls a closed orbit.

Again with the list of curvature values we compute

$$R(e_1, \gamma')\gamma' = (-\frac{m}{G^3} - \omega^2)e_1 =: K_\rho e_1.$$

Since we have two eigenvectors $e_1, e_3$ of $X \mapsto R(X, \gamma')\gamma'$ also $f_2$ is an eigenvector, and since $Ric(\gamma') = 0$ the sum of the three eigenvalues is zero. Hence

$$R(f_2, \gamma')\gamma' = \frac{m}{G^3} f_2 =: K_\varphi f_2.$$

Do we get interesting Jacobi fields from this information? First some trivial ones: $f_2$ is the restriction of a Killing field, it describes the neighboring geodesics obtained by adding a constant to the parameter $s$. One can also check that

$$J(s) := e_1(s) + \frac{K_\rho - \omega^2}{2\omega}(s - s_0) f_2(s)$$

are Jacobi fields. They describe concentric circular orbits of different radius and since they have different orbit velocity the non-radial $f_2$-component is needed.

We have not yet all Jacobi fields in span$\{e_1, f_2\}$ and therefore we try the Ansatz

$$\begin{aligned}
&J(s) := \lambda(s) f_2(s) + \mu(s) e_1(s). \\
&(\lambda f_2 + \mu e_1)'' + R(\lambda f_2 + \mu e_1, \gamma')\gamma' = 0 \qquad \text{gives the ODEs:} \\
&(\lambda' + 2\omega\mu)' = 0, \qquad\quad -2\omega\lambda' + \mu'' + (K_\rho - \omega^2)\mu = 0.
\end{aligned}$$

We put the constant function $\lambda' + 2\omega\mu = const$ in the second equation:

$$\mu'' + (K_\rho + 3K_\varphi)\mu = const \cdot 2\omega$$

and since $\mu = const_1$ was already discussed for the concentric orbits we are left with

$$0 = \mu'' + (K_\rho + 3K_\varphi)\mu = \mu'' + \omega^2(1 - \frac{6m}{G})\mu.$$

This ODE first gives the desired perihelion advance since the frequency of these Jacobi fields is by the factor $\sqrt{1 - \frac{6m}{G}} \approx 1 - \frac{3m}{G_{Mercury}} = 1 - 4.5/(5.8 \cdot 10^7) = 1 - 0.77 \cdot 10^{-7}$ smaller than the orbit frequency $\omega$, $T_{Mercury} = 88$ days. This says: the relativistic perihelion advance is in 88 days by $360° \cdot 0.77 \cdot 10^{-7} = 0.1''$ (plus $1.28''$ classical contribution), or in a century by $36500/88 \cdot 0.1 = 41.4''$.
But secondly, this argument works only if $G > 6m$, for $G < 6m$ there are exponentially growing and exponentially decaying Jacobi fields. This says that small orbit perturbations can blow up exponentially: these circular orbits are unstable.

One word about *relativistic corrections.* The curvature values we have worked with contain the term $m/G^3$ and various correction factors $(1 - const\, m/G)$. If we ignore these correction factors our computations give the Newtonian predictions. Therefore the term *relativistic corrections* refers to by how much the factors $(1 - const\, m/G)$ change the result.

## Spinning Planets

A solid body is not part of Relativity Theory, because an attempt to accelerate a really solid and extended object does not agree with the simultaneity discussion of Special Relativity. Also, I am unaware of rotating "solid" objects where the orbit speeds are comparable to the velocity of light. Still, for slowly rotating almost solid objects like planets the question remains whether the Schwarzschild geometry predicts a different behaviour than Newtonian theory. By viewing a planet as a *test particle with tensor of inertia* we treat a classical object in a relativistic geometry.

We describe how the curvature tensor of space time acts on a rotating solid object with tensor of inertia $\Theta$. The center of mass has the world line $\gamma(s)$ ($s$ is proper time) and the separation vectors $X$ describe the mass points relative to the center. We say that the object rotates with angular velocity $\vec{\omega} = (\omega_1, \omega_2, \omega_3)$ if the velocities of all the mass points are obtained as

$$X'_N(s) = \vec{\omega}(s) \times X_N(s) = \begin{pmatrix} 0 & \omega_3 & -\omega_2 \\ -\omega_3 & 0 & \omega_1 \\ \omega_2 & -\omega_1 & 0 \end{pmatrix}(s) \cdot X_N(s).$$

The definitions of the angular momentum and the tensor of inertia are

**Angular Momentum** $L := \sum_N m_N X_N \times X'_N = \sum_N m_N X_N \times (\vec{\omega} \times X_N)$

$$= \int X_m \times (\vec{\omega} \times X_m) dm$$

**Tensor of Inertia** $\Theta : \vec{\omega} \to L = \Theta(\vec{\omega}) = \int X_m \times (\vec{\omega} \times X_m) dm.$

Note: $$X \times (\vec{\omega} \times X) = \begin{pmatrix} x_2^2 + x_3^2 & -x_1 x_2 & -x_1 x_3 \\ -x_1 x_2 & x_1^2 + x_3^2 & -x_2 x_3 \\ -x_1 x_3 & -x_2 x_3 & x_1^2 + x_2^2 \end{pmatrix} \cdot \vec{\omega}.$$

The last matrix (or alternatively $\langle \vec{a}, \vec{b} \times \vec{c}\, \rangle = \det((\vec{a}, \vec{b}, \vec{c})\,)$ shows that $\Theta$ is a symmetric map and therefore has a basis of eigenvectors $\{\vec{e}_1, \vec{e}_2, \vec{e}_3\}$ in which the above matrix simplifies to

$$\begin{pmatrix} \Theta_1 & 0 & 0 \\ 0 & \Theta_2 & 0 \\ 0 & 0 & \Theta_3 \end{pmatrix} = \int \begin{pmatrix} x_2^2 + x_3^2 & 0 & 0 \\ 0 & x_1^2 + x_3^2 & 0 \\ 0 & 0 & x_1^2 + x_2^2 \end{pmatrix} dm$$

$$\operatorname{trace} \Theta = 2 \int |X_m|^2 dm, \qquad \int x_j^2 dm = \frac{1}{2} \operatorname{trace} \Theta - \Theta_j.$$

The rotational position of the body will *always* be specified by giving $\{\vec{e}_1(s), \vec{e}_2(s), \vec{e}_3(s)\}$, the *moving body-frame*.

In the absence of exterior moments the angular momentum is constant. In the case of a nonvanishing curvature tensor we have the time dependent acceleration
$X''(s) = -R(X,\gamma')\gamma' =: -R_{\gamma'}(X)$. A non-vanishing torque (or moment) results:

$$M(s) = -\sum_N m_N X_N(s) \times R_{\gamma'}(X_N)(s) = -\int X_m(s) \times R_{\gamma'(s)}(X_m(s))\, dm.$$

The rotational behaviour of the solid body along the world line $\gamma$ is governed by the ODE

$$\frac{D}{ds}L(s) = M(s).$$

Note that this description is not very different from the classical treatment: Instead of our symmetric curvature tensor contribution $X \mapsto R_{\gamma'}(X)$ one has the inhomogeneous gravitational field $\operatorname{grad}\Phi$ of the sun and the torque arises from the difference of the field at the planet's center, $X = 0$, and at the mass points $X_m$. Since $X_m$ is assumed small the torque is given by $X_m \times \operatorname{Hess}\Phi(X_m)$. The Hessian is symmetric and the three eigenvalues are $+m/G^3$, $+m/G^3$, $-2m/G^3$, the same as for $R_{\gamma'}$ – except for the *relativistic correction factors* $(1 - const\ m/G)$.

The goal is to discuss the above ODE in terms of $\Theta$ and $R_{\gamma'}$, without reference to the individual mass points. For **vanishing curvature** we have the following treatment by Euler. Everything is expressed in the moving body-frame $\{\vec{e}_1(s), \vec{e}_2(s), \vec{e}_3(s)\}$.

$$\begin{aligned}
\vec{\omega}(s) &= \sum_i \omega_i(s)\vec{e}_i(s) \quad \text{with} \quad \frac{D}{ds}\vec{e}_i = \vec{\omega}(s) \times \vec{e}_i(s) \\
L(s) &= \Theta(s) \cdot \vec{\omega}(s) = \sum_i \omega_i(s)\Theta_i\vec{e}_i(s) \\
\frac{D}{ds}L(s) &= \sum_i \omega_i'(s)\Theta_i\vec{e}_i(s) + \sum_i \omega_i(s)\Theta_i(\vec{\omega} \times \vec{e}_i(s)) \\
&= \sum_i \omega_i'(s)\Theta_i\vec{e}_i(s) + \vec{\omega}(s) \times L(s). \\
&= \sum_i \omega_i'(s)\Theta_i\vec{e}_i(s) + \sum_i \left( \begin{pmatrix} \omega_1(s) \\ \omega_2(s) \\ \omega_3(s) \end{pmatrix} \times \begin{pmatrix} \Theta_1\omega_1(s) \\ \Theta_2\omega_2(s) \\ \Theta_3\omega_3(s) \end{pmatrix} \right)_i \vec{e}_i(s)
\end{aligned}$$

Therefore one first has to solve Euler's first order ODE for the $\omega_i(s)$:

$$\begin{pmatrix} \Theta_1\omega_1(s) \\ \Theta_2\omega_2(s) \\ \Theta_3\omega_3(s) \end{pmatrix}' + \begin{pmatrix} \omega_1(s) \\ \omega_2(s) \\ \omega_3(s) \end{pmatrix} \times \begin{pmatrix} \Theta_1\omega_1(s) \\ \Theta_2\omega_2(s) \\ \Theta_3\omega_3(s) \end{pmatrix} = 0$$

and then use these $\omega_i(s)$ to integrate $\frac{D}{ds}\vec{e}_i = (\sum_i \omega_i(s)\vec{e}_i(s)) \times \vec{e}_i(s)$.

**Non-vanishing curvature.** If $R_{\gamma'} \neq 0$ the system does not reduce to two first order integrations, the system stays coupled. But otherwise the strategy is the same, except for the initial obstacle that the torque is still expressed via the action of $R_{\gamma'}$ on all the mass points and not via $\Theta$ and $R_{\gamma'}$. Expand the moment integral with respect to the eigen frame $\{\vec{e}_1(s), \vec{e}_2(s), \vec{e}_3(s)\}$ of $\Theta$.

$$\begin{aligned} M(s) &= -\int X_m(s) \times R_{\gamma'(s)}(X_m(s))\, dm = -\sum_{i,j} \vec{e}_i \times R_{\gamma'}(\vec{e}_j) \int x_i x_j dm \\ &= -\sum_i \vec{e}_i \times R_{\gamma'}(\vec{e}_i) \int x_i^2 dm \quad \text{(mixed terms vanish in eigen basis)} \\ &= -\sum_i \vec{e}_i \times R_{\gamma'}(\vec{e}_i)(\frac{1}{2}\operatorname{trace}\Theta - \Theta_i) = \sum_i \vec{e}_i \times R_{\gamma'}(\vec{e}_i)\Theta_i. \end{aligned}$$

We used here and in the next line that for symmetric $A_{i,j}$ : $\quad \sum_{i,j} \vec{e}_i \times A_{i,j}\vec{e}_j = 0$.

$$M(s) = \sum_i \vec{e}_i \times R_{\gamma'}(\Theta(\vec{e}_i)) = \frac{1}{2}\sum_i \vec{e}_i \times [R_{\gamma'}, \Theta](\vec{e}_i).$$

In particular: $\qquad M(s) = 0 \quad \text{if} \quad \Theta_1 = \Theta_2 = \Theta_3.$

Here we have obtained the result that the axis of a perfectly symmetric rotating solid body along a world line $\gamma$ is described by a *parallel* vector field in the normal bundle of $\gamma$. In the physics literature this is called *Fermi-Walker transport.*

Finally we rewrite the moment as linear combination of the eigenvectors so that the result can be combined with the computation for zero curvature. Note that the Lorentz metric $g$ is Riemannian in the normal spaces of $\gamma$.

$$\begin{aligned} M(s) &= \frac{1}{2}\sum_{i,k} g\left(\vec{e}_i \times [R_{\gamma'}, \Theta](\vec{e}_i), \vec{e}_k\right) \cdot \vec{e}_k = \frac{1}{2}\sum_{i,k} \det\left(\vec{e}_k \times \vec{e}_i, [R_{\gamma'}, \Theta](\vec{e}_i)\right) \cdot \vec{e}_k \\ &= \sum_{(i,j,k)=(1,2,3)} g\left(\vec{e}_j, [R_{\gamma'}, \Theta](\vec{e}_i)\right) \cdot \vec{e}_k \qquad \text{(cyclic sum)} \\ &= \sum_{(i,j,k)} (\Theta_i - \Theta_j) g\left(\vec{e}_j, R_{\gamma'}(\vec{e}_i)\right) \cdot \vec{e}_k =: \sum_{(i,j,k)} (\Theta_i - \Theta_j) R_{i,j}\vec{e}_k \end{aligned}$$

This leaves us with the Euler equations coupled to the curvature

$$\begin{pmatrix} \Theta_1\omega_1(s) \\ \Theta_2\omega_2(s) \\ \Theta_3\omega_3(s) \end{pmatrix}' + \begin{pmatrix} \omega_1(s) \\ \omega_2(s) \\ \omega_3(s) \end{pmatrix} \times \begin{pmatrix} \Theta_1\omega_1(s) \\ \Theta_2\omega_2(s) \\ \Theta_3\omega_3(s) \end{pmatrix} = \begin{pmatrix} (\Theta_2 - \Theta_3)R_{2,3} \\ (\Theta_3 - \Theta_1)R_{3,1} \\ (\Theta_1 - \Theta_2)R_{1,2} \end{pmatrix},$$

but to get the curvature components $R_{i,j} = g(\vec{e}_i, R_{\gamma'}(\vec{e}_j))$ one has simultaneously to integrate

$$\frac{D}{ds}\vec{e}_i = \left(\sum_i \omega_i(s)\vec{e}_i(s)\right) \times \vec{e}_i(s). \qquad \text{Q.E.D.}$$

## The Kruskal Extension: beyond 2m

People were puzzled by the *Schwarzschild singularity* at $G = 2m$ for over forty years until Kruskal found an analytic extension of the Schwarzschild geometry. This did not answer all natural questions since no information from that other part can reach us and therefore no hypotheses can be checked. Those forty years of puzzlement suggest that the Kruskal extension is not at all obvious. However I learnt from the book *Allgemeine Relativitätstheorie* by *Hans Stephani* the following beautiful derivation.

We noticed that the classical coordinates for the Schwarzschild geometry are tuned to the Killing observers and since their world line acceleration becomes infinite at $G = 2m$ these coordinates *cannot* work across that limit. So let us look for another family of observers with the goal of choosing good coordinates for them! Candidates are particles that fall in radially from infinity, starting with limit velocity zero. The section on falling particles gives:

$$F = G' = (1 - \frac{2m}{G} - \frac{\Lambda}{3}G^2)^{1/2} \qquad \text{(definition of metric)}$$
$$F^2(\rho(s))t'(s) = T = 1 \qquad \text{(being at rest at infinity)}$$
$$G^2(\rho(s))\varphi'(s) = \Omega = 0,\ \varphi(s) = const \qquad \text{(falling radially)}$$
$$\rho'(s)^2 = \frac{T^2}{F^2(\rho)} - 1 = \frac{1}{F^2(\rho)} - 1 = (\frac{2m}{G} + \frac{\Lambda}{3}G^2)(1 - \frac{2m}{G} - \frac{\Lambda}{3}G^2)^{-1}$$

This leads to two different expressions for the proper time on these world lines $\gamma$.

$$ds_1 = F^2(\rho)dt(\gamma') = (1 - \frac{2m}{G} - \frac{\Lambda}{3}G^2)dt(\gamma')$$
$$ds_2 = -(\frac{1}{F^2} - 1)^{-1/2}d\rho(\gamma') = -(1 - F^2)^{-1/2} \cdot dG(\gamma').$$

Any convex combination of $ds_1$ and $ds_2$ will still compute proper time on our world lines. The key is to ask, whether such a combination can be found that is the *differential of a function*! One does not even have to solve equations because it is so easy to guess:

$$ds = \frac{1}{F^2}ds_1 + (1 - \frac{1}{F^2})ds_2 = dt(\gamma') + \frac{\sqrt{1 - F^2}}{F^2}\,dG(\gamma')$$

We define two new functions of $t$ and $G$ (use $F(x) := (1 - \frac{2m}{x} - \frac{\Lambda}{3}x^2)^{1/2}$ )

$$S := t + \int_{3m}^{G} \frac{\sqrt{1 - F^2}}{F^2}(x)\,dx \quad \text{so that: } ds = dS(\gamma')$$
$$R := S + \int_0^G \frac{dx}{\sqrt{1 - F^2(x)}} \quad \text{so that: } dR = dS + \frac{dG}{\sqrt{1 - F^2}}, \quad dR(\gamma') = 0.$$

Note that $R - S$ is a strictly monoton function of $G$ and therefore can be inverted as $G = G(R - S)$, explicitly if $\Lambda = 0$. Then $t$ can be computed from $S$ and $G$. Therefore

$(G,t) \mapsto (R,S)$ is a coordinate change adapted to the inward falling particles. The function $R$ is not as natural as $S$, we chose for $R$ a combination of $S, G$ that is constant on our guiding world lines. This coordinate change is successful, the metric expression does not develop singularities for positive values of the function $G = G(R - S)$,
We claim:

$$\textbf{Schwarzschild:}\quad \frac{dG^2}{1 - 2m/G - \frac{\Lambda}{3}G^2} + G^2 d\sigma^2 - (1 - \frac{2m}{G} - \frac{\Lambda}{3}G^2)dt^2$$

$$\textbf{Kruskal:}\quad = (\frac{2m}{G} + \frac{\Lambda}{3}G^2)dR^2 + G^2 d\sigma^2 - dS^2.$$

Note that $G : (0,\infty) \to (0,\infty)$ is a strictly monoton function of $(R - S)$.
The range of these new coordinates is the half plane $-\infty < S < R < \infty$.

Proof:

$$F^2 dt^2 = \left(-\frac{\sqrt{1-F^2}}{F} \cdot dG + F \cdot dS\right)^2$$

$$(1-F^2)dR^2 = \left(\sqrt{1-F^2} \cdot dS + dG\right)^2$$

$$F^2 dt^2 + (1-F^2)dR^2 = \frac{1}{F^2}dG^2 + dS^2 + 0 \cdot dG\,dS$$

$$(1-F^2)dR^2 - dS^2 = \frac{dG^2}{F^2} - F^2 dt^2.$$

$$\text{Recall}\quad F^2 = (1 - \frac{2m}{G} - \frac{\Lambda}{3}G^2) \qquad \text{Q.E.D.}$$

Near the boundary of the $R$-$S$-halfplane, as $(R - S) \to 0$, the curvature values $m/G^3$ blow up. Along this boundary of the Kruskal coordinates we have curvature singularities and therefore no further extension is possible.
In these new coordinates the rotational symmetries are as before. The time translation is now $(R,\ S) \to (S + c,\ R + c)$ which indeed leaves the function $G$ unchanged, the corresponding Killing field is $(1,0,1)$ of timelike length squared $\frac{2m}{G} + \frac{\Lambda}{3}G^2 - 1$, as before. At the two positive values of $G$ where the Schwarzschild form of the metric becomes degenerate, the Killing field has lightlike values and becomes even spacelike beyond those points. The geodesic equation of falling particles can be discussed with the same conserved quantities: let $\gamma(s) = (R(s), \sigma(s), S(s))$, then we get as before a first order ODE for $R(s)$:

$$(\frac{2m}{G} + \frac{\Lambda}{3}G^2)R'^2 + G^2\sigma'^2 - S'^2 = -1, \quad G^2\sigma' = \Omega, \quad (\frac{2m}{G} + \frac{\Lambda}{3}G^2)R' - S' = T,$$

and $R' = S'$ for circular orbits.

## The Kerr Solution: Frame Dragging

A drawback of the Schwarzschild geometry is its spherical symmetry. Models of star generation suggest that stars should rotate and therefore black holes should also have angular momentum. In 1963 R. Kerr found a solution with rotational symmetry without solving a PDE, quite surprisingly his solution is in terms of 1-dimensional simple functions. The most exciting feature of this solution is that planets whose orbital angular momentum is parallel respectively anti-parallel to the angular momentum of the star have *different* periods. One says, the rotating star drags space along. This *frame dragging* was predicted by *Lense and Thirring* before 1920.

In spite of the simple functions in terms of which the Kerr solution is written, it is geometrically a complicated space. I believe it gives more insight to study it with the help of symbolic and numerical computer programs than to do this by hand. With one exception: people have been particularly interested in planets orbiting the star in its equatorial plane. This requires only to look at a 3-dimensional totally geodesic subspace of the Kerr geometry (set the polar angle $\theta = \pi/2$), and this restriction is not too much more complicated than the Schwarzschild geometry. We visualize it by using polar coordinates in a horizontal plane and take the $t$-axis vertical. The vertical lines $t \to (r_0, \varphi_0, t)$ are the orbits of the isometric time translation and will be called (as in the Schwarzschild case) world lines of Killing observers. The horizontal planes $t = const$, with a hole in the middle, are no longer totally geodesic (time reflection $t \mapsto -t$ is not a Kerr isometry), nor are they orthogonal to the Killing world lines. Everything moving on circles around the star has world lines on the concentric cylinders $r = const.$. These 2-dimensional cylinders have an $r$-dependent flat metric, its intrinsic geodesics are straight lines (rolled onto the cylinder), the time like ones of these are the world lines of objects that circle the star with constant velocity. The planetary or photon orbits are also *extrinsic* geodesics and we want to find these without computing the Christoffel symbols. An intrinsic geodesic is an extrinsic geodesic if the normal variation ($r$-direction) does not change the length in first order.

The metric of the 3-dim. equator section of the Kerr geometry is:

$$ds^2 = \frac{r^2}{r^2 - 2mr + a^2} dr^2 + (r^2 + a^2(1 + \frac{2m}{r}))d\varphi^2 - \frac{4am}{r} d\varphi\, dt - (1 - \frac{2m}{r})dt^2,$$

$$g(\begin{pmatrix} 0 \\ x \\ y \end{pmatrix}, \begin{pmatrix} 0 \\ x \\ y \end{pmatrix}) = (r^2 + a^2(1 + \frac{2m}{r})) \cdot x^2 - \frac{8am}{r} xy - (1 - \frac{2m}{r}) \cdot y^2 =: f_{xy}(r).$$

$$\frac{d}{dr} f_{xy}(r) = \frac{2m}{r^2} y^2 - \frac{8am}{r^2} xy - (2r - \frac{2ma^2}{r^2})x^2 \overset{(!)}{=} 0 \quad \text{hence}$$

$$\frac{y}{x} = \pm 2a + \sqrt{\frac{r^3}{m}} \sqrt{1 + \frac{3ma^2}{r^3}} = \sqrt{\frac{r^3}{m} + 3a^2} \pm 2a =: q_\pm(r, a)$$

are the extrinsic geodesic directions on the cylinders. (For surfaces:*asymptote directions*)

The tangent vectors of planetary world lines at radial coordinate $r$ therefore are:

$$\gamma' = (0, \pm\omega, \pm\omega \cdot q_\pm(r, a))$$

For photons one wants to determine $\omega$ such that $g(\gamma', \gamma') = 0$, for particles one wants $g(\gamma', \gamma') = -1$. (Such $\omega$ may not exist, e.g. for $r^3/m = 3a^2$ one $q_-(r, a) = 0$ and $(0, -\omega, 0)$ is space like.) After one has determined $\omega$, then the orbit $\gamma(s)$ has completed one revolution – as observed by the Killing observer – if $\omega \cdot s = \pm 2\pi$, the coordinate time for one revolution (observed by the Killing guy) therefore is

$$\text{Koordinate Period } T_+ = 2\pi \cdot q_+(r, a) \neq 2\pi \cdot q_-(r, a) = \text{Koordinate Period } T_-.$$

This is the first important result: *For the same radial coordinate r the period time depends on whether the planet circles such that its orbit angular momentum is parallel* $(+)$ *or anti-parallel* $(-)$ *to the spin of the Kerr geometry. The difference is:* $T_+ - T_- = 8\pi a$.

Next we look for photon orbits, i.e., for given $m, a$ find $r$ such that $\gamma' = \pm(0, \omega, q_\pm(r, a) \cdot \omega)$ is a null direction. I could not solve this explicitly, but I could find the pair $r/m,\ a/m$ in terms of an auxiliary parameter $\lambda$ by writing

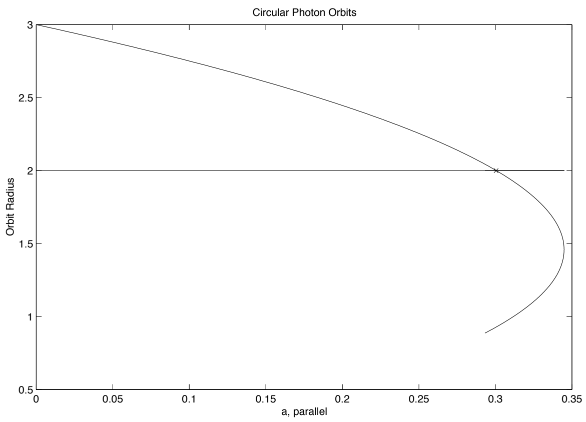
$$\frac{r^3}{m} = (\frac{1}{\lambda^2} - 3)a^2, \quad \text{hence} \quad \frac{y}{x} = q_\pm(r, a) = a(\frac{1}{\lambda} \pm 2).$$

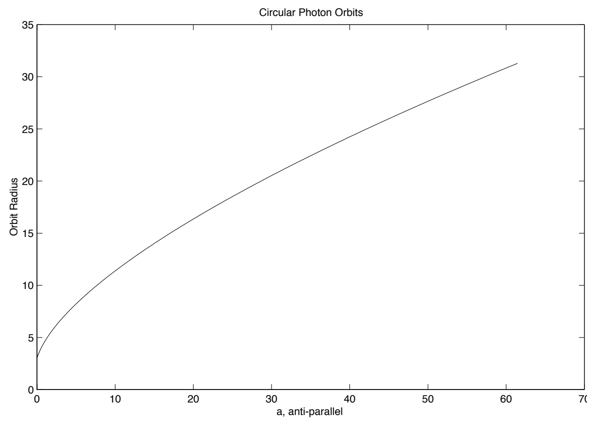
Instead of writing $\pm 2$ I use $\lambda > 0$ for orbits with parallel angular momentum and $\lambda < 0$ for orbits with anti-parallel angular momentum and always $q_\pm(r, a) = a|2 + 1/\lambda|$. Now $g(\gamma', \gamma')$ with $\gamma' = (0, \pm 1, q_\pm(r, a))$ is reasonably simple:

$$g(\gamma', \gamma') = 3r^2 - a^2(3 + \frac{1}{\lambda^2} + \frac{4}{\lambda}) \overset{(!)}{=} 0.$$

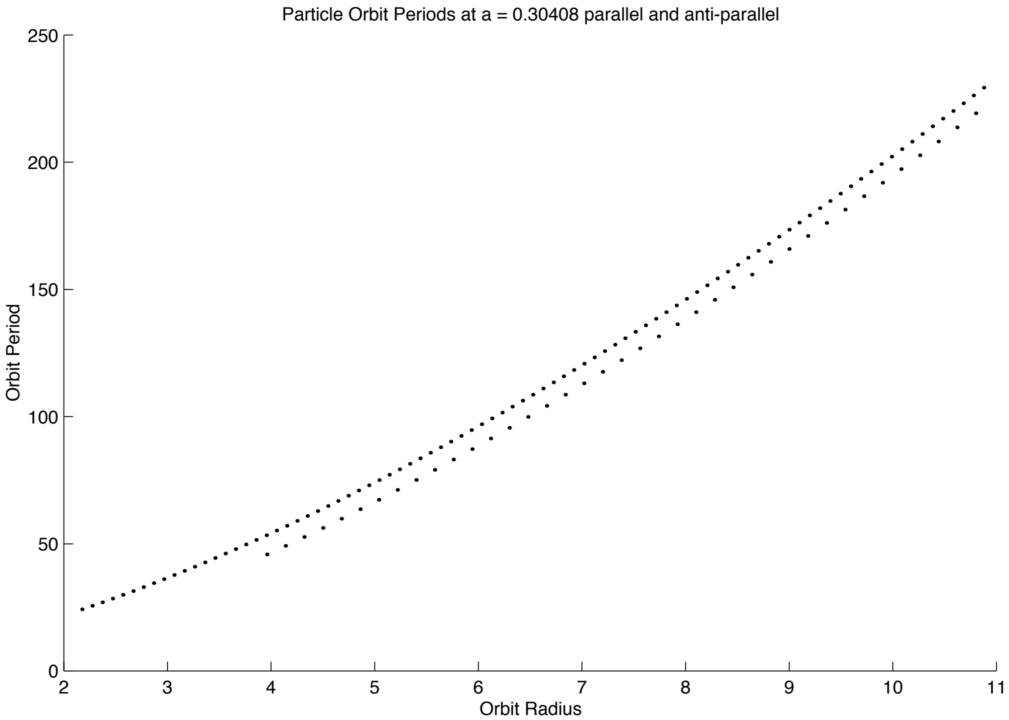
From this equation and the previous one (the definition of $\lambda$) we eliminate $a^2$ to obtain $r/m$ in terms of $\lambda$ and then also $a/m$ in terms of (positive or negative) $\lambda$:

$$\frac{r}{m} = 3\frac{1 - 3\lambda^2}{1 + 3\lambda^2 + 4\lambda}, \qquad \frac{a}{m} = \frac{r}{m}|\lambda|\sqrt{\frac{3}{1 + 3\lambda^2 + 4\lambda}}.$$




At $\lambda = 0$ we obtain the Schwarzschild case: $r = 3m,\ a = 0$. For positive $\lambda$ note that $r/m$ is monotone decreasing with $\lambda$ and drops below $r = 2m$ (where the present form of the Kerr metric is no longer defined) already at $a = 0.3m, \lambda \approx 0.1044$. On the other hand, for negative $\lambda$ one finds that $r/m$ is a monotone increasing function of $a/m$ (for example at $a/m = 61$ we have $r/m = 31$). The main point is that photons circling parallel to the black hole's spin behave very different from photons circling in the opposite direction. Another simple consequence is this: since the null directions on the cylinder $r = const$ separate the space like directions from the time like ones, and since $q_\pm(r, a)$ is monotone increasing in $r$ there can be no circular particle orbits on radii smaller than the photon orbits. For a value of $a = 0.304$ we compare the periods of circular orbits in the following diagram, the anti-parallel ones take less time ($T_+ - T_- = 8\pi a$):




Next we wish to look at the differential equation of orbits (of particles or photons) on which the radial coordinate is not constant.

With the following abbreviations

$$f_{11}(r) := \big(1 - \frac{2m}{r} + \frac{a^2}{r^2}\big)^{-1}, \quad f_{22}(r) := r^2 + a^2 + \frac{2ma^2}{r},$$
$$f_{23}(r) := \frac{-4ma}{r}, \quad f_{33}(r) := -1 + \frac{2m}{r}$$

we get the metric in the form

$$ds^2 = f_{11}(r)dr^2 + f_{22}(r)d\varphi^2 + f_{23}(r)d\varphi dt + f_{33}(r)dt^2.$$

The Killing fields $(0, 1, 0)$ and $(0, 0, 1)$ give two constants $\Omega, T$ of the motion

$$g((r', \varphi', t'), (0, 1, 0)\ ) = f_{22}(r)\varphi' + f_{23}(r)t' =: \Omega,$$
$$g((r', \varphi', t'), (0, 0, 1)\ ) = f_{23}(r)\varphi' + f_{33}(r)t' =: T.$$
$$g((r', \varphi', t'), (\rho', \varphi', t')\ ) = f_{11}(r){r'}^2 + \Omega\varphi' + Tt'.$$

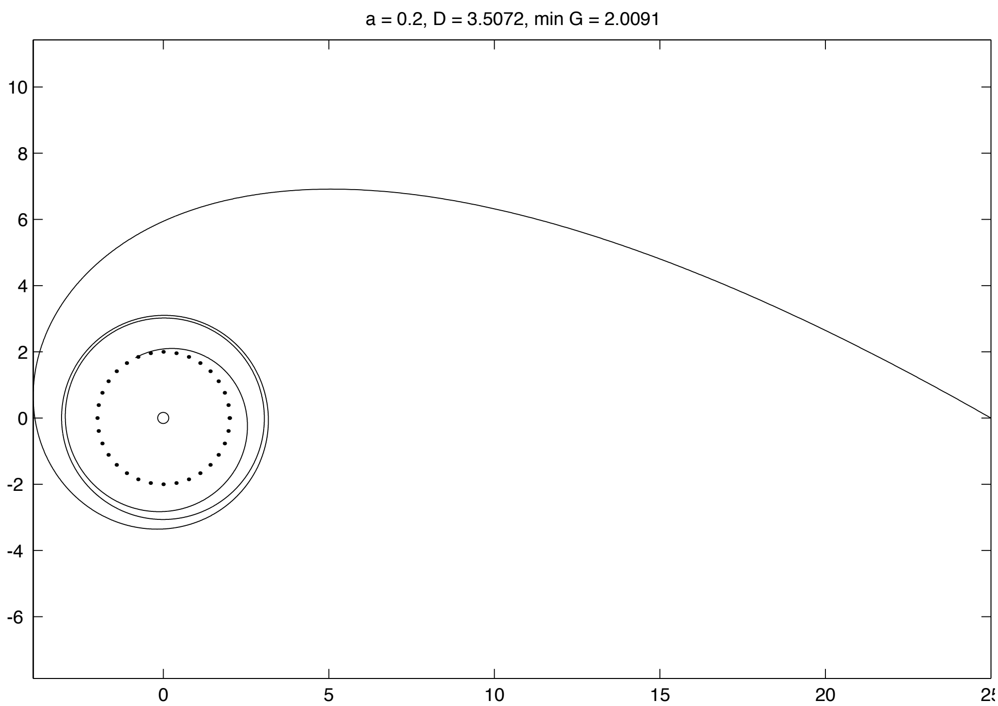
The matrix

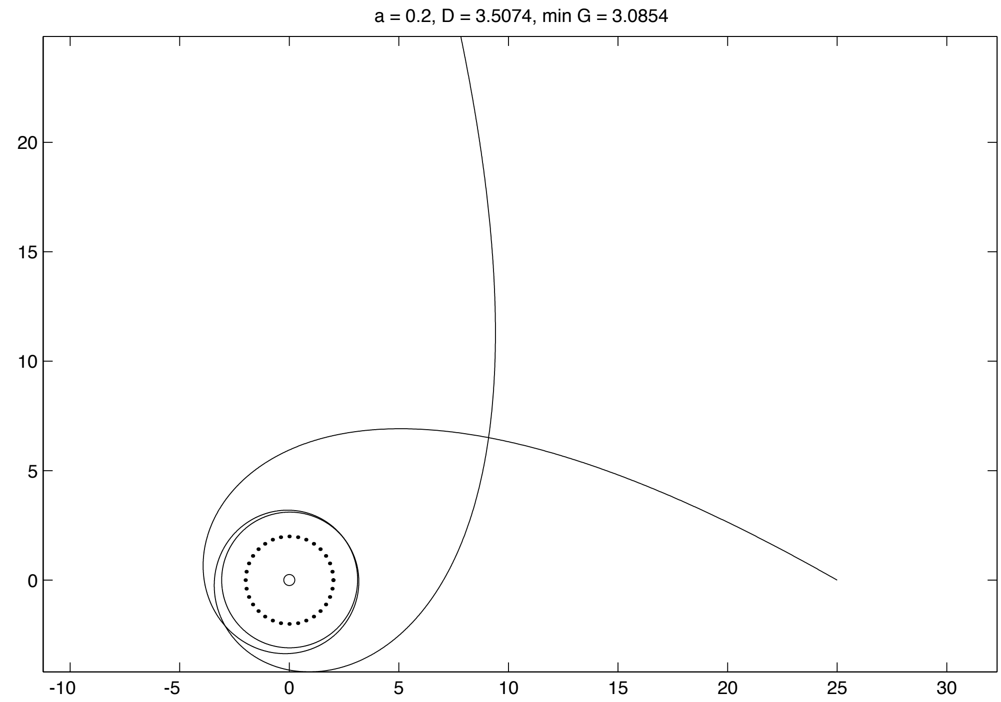
$$M := \begin{pmatrix} f_{22}(r) & f_{23}(r) \\ f_{23}(r) & f_{33}(r) \end{pmatrix}, \quad M \cdot \begin{pmatrix} \varphi' \\ t' \end{pmatrix} = \begin{pmatrix} \Omega \\ T \end{pmatrix}$$

has non-vanishing negative determinant where $\{r > 2m\}$ and therefore gives the first order ODE for the radial part $r(s)$ of the orbit:

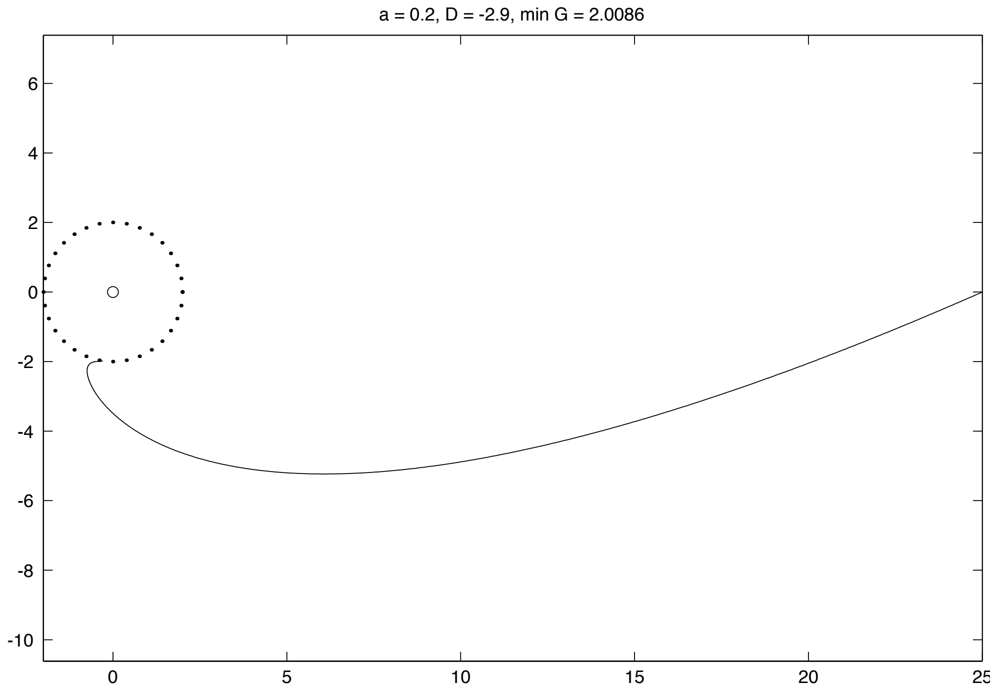
$$f_{11}(r){r'}^2 + (\Omega, T) \cdot M^{-1}(r) \cdot \begin{pmatrix} \Omega \\ T \end{pmatrix} = 0 \text{ (for photons), respectively } = -1 \text{ (for particles).}$$

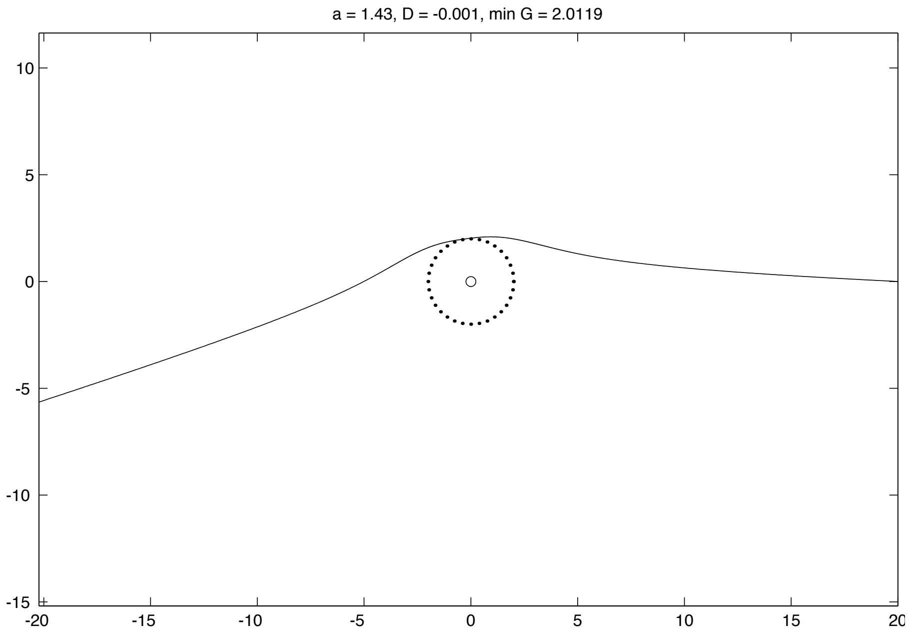
In this generality this is very similar to the Schwarzschild case. However, the detailed behavior of the solutions is much more complicated and I only know how to do that numerically. For photons we may assume $T = 1$ and discuss the solutions in their dependence on $\Omega$. For sufficiently large $|\Omega|$ photons falling in from large values of $r$ will approach the star but then $r'$ changes sign and they come out again. For sufficiently small $|\Omega|$ the photons will fall into the black hole. For **negative** $\Omega$ one needs considerably larger $|\Omega|$ for the photon **not** to be captured. For larger values of $a$, captured photons have orbits which, on their last stretch before they reach $r = 2m$, fall in almost radially. For small values of $a$ one notices a sign change in $\varphi'$, on the last piece of their orbit captured photons move in the direction parallel to the spin of the black hole. – For **positive** $\Omega$ considerably smaller values of $\Omega$ are sufficient for the photon to escape the black hole. But a capture minimum remains, for example in the case $m = 1,\ T = 1$ photons with $\Omega = 1$ will be captured, no matter what the spin $a$ of the black hole is. I omit pictures because the behaviour of *particles* is more dramatic than that of photons.
The first three diagrams have small angular momentum of the star, $a = 0.2$, the first two have positive (and only slightly different) orbit angular momentum, the first one shows capture:



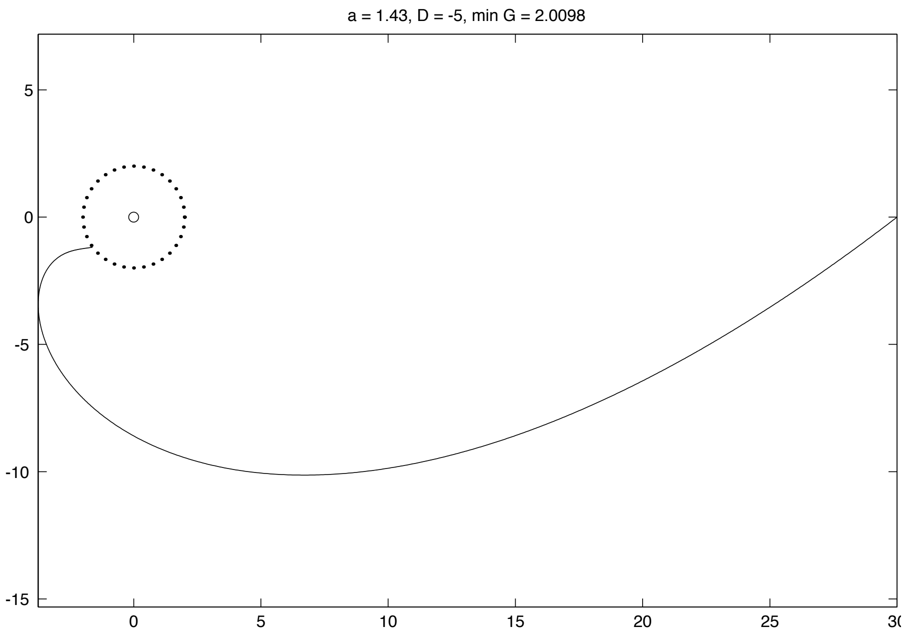


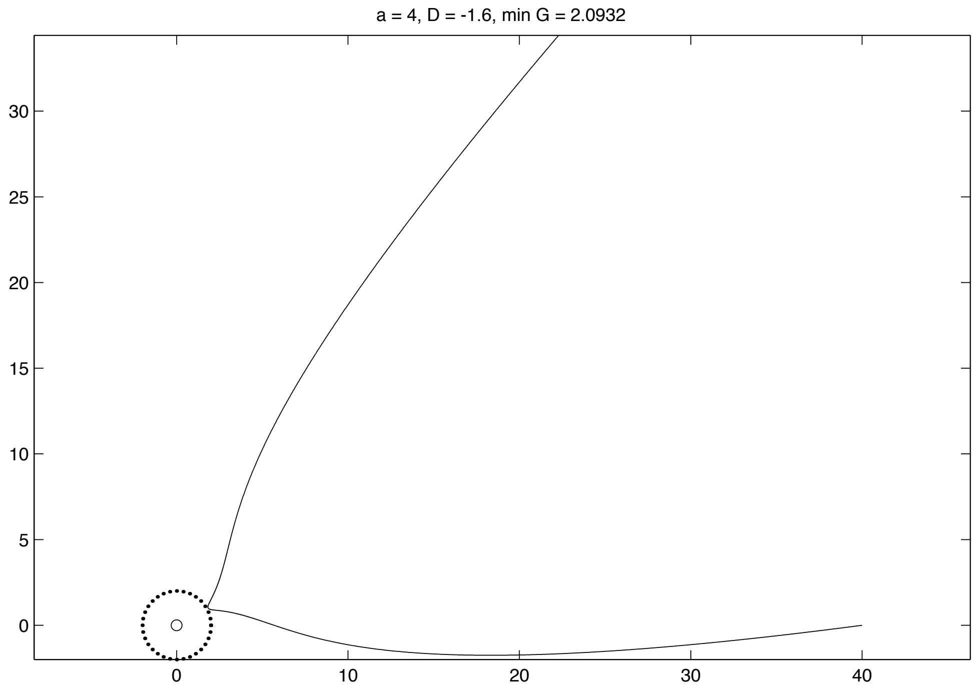

In the second diagram the particle can leave the gravitational field, but not before going around three times. This can also happen in the Schwarzschild geometry, but not in a Newtonian system. – Below we have a *negative* orbit momentum. Near $\{r = 2m\}$ the field of the star forces the particle around into the positive direction, then it is captured.



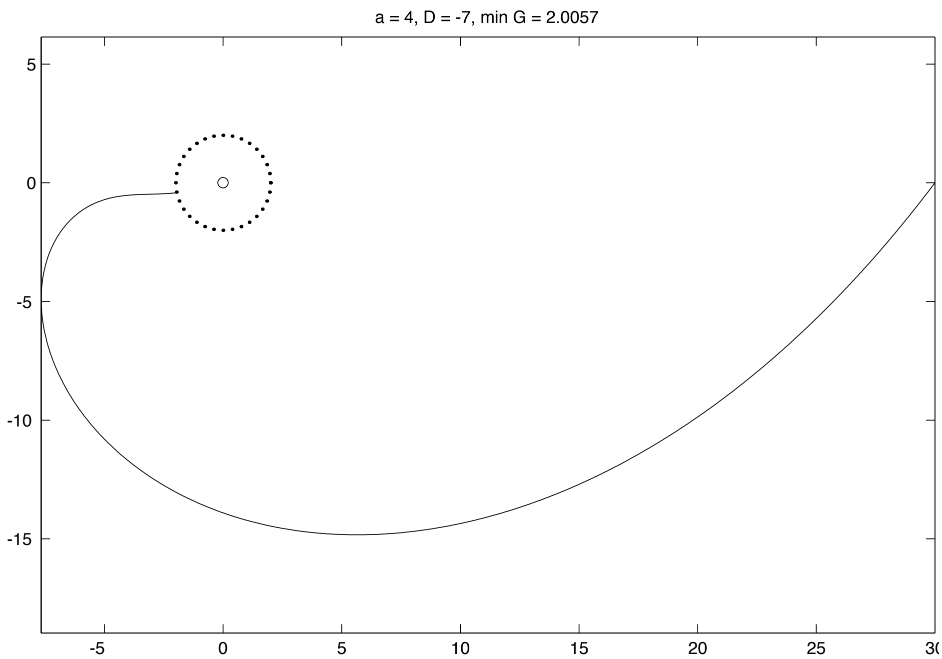



In these two diagrams the star has angular momentum $a = 1.43$. About this size of the angular momentum is critical in that *no particle with positive orbit momentum* can be captured by the star. In the diagram above the particle has small negative orbit momentum, but is still not captured. The diagram below shows the turning around of particles with larger *negative* orbit momentum near $\{r = 2m\}$ and subsequent capture.






These two diagrams show orbits near a star with angular momentum $a = 4$. It is assumed that black holes cannot have so large an angular momentum, a) because an imploding star could not have rotated fast enough before its implosion and b) because no angular momentum can be added to an existing black hole beyond the limit $a/m > 1.43$, since no particle with positive orbit momentum can be captured.

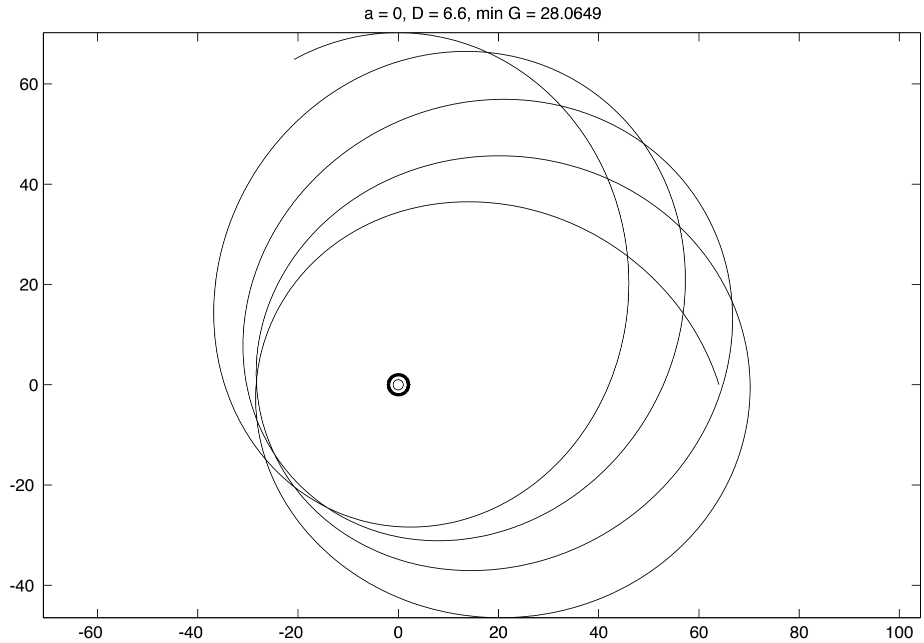


Because the fact, that angular momentum of a Kerr solution cannot be increased by throwing orbiting particles with positive orbital momentum into the black hole, is very important, we check our numerical computation against the above formulas: On an orbit with $\Omega = 0$ we need $r' = 0$ to happen just outside $\{r = 2m\}$. We consider orbits that are at rest at infinity, i.e., $T = -1$. We ask: what is the smallest $a$ so that this happens? The ODE gives:

$$0 + T^2 f_{22}(r)/\det(M)\big|_{r=2m} = -1 \quad \Rightarrow \quad 1 = \frac{4a^2}{f_{22}(2m)} = \frac{4a^2}{4m^2 + 2a^2}$$
$$\Rightarrow \quad a = \sqrt{2} \cdot m.$$

This is indeed the numerically discovered critical value.

Finally, since our orbit computations can also show the *perihelion advance* we give one such example:



## Stress Energy Tensor

Simple Examples and Geometric Consequences, a Schur Theorem

Notational conventions. For the Ricci Tensor I will use different names for its bilinear version: $ric(v,w)$ and its 1-1-Tensor version: $Ric(v)$, and of course: $ric(v,w) = g(Ric(v),\ w)$. The divergence free part of the Ricci Tensor is the Einstein Tensor $G$:

$$G := Ric - \frac{1}{2}(\mathrm{trace}\, Ric) \cdot \mathrm{id}\,, \qquad \mathrm{trace}\, G = -\mathrm{trace}\, Ric.$$

**The Einstein Equation**

$$8\pi T = G + \Lambda \mathrm{id}\,.$$

Of course, without further words this means nothing: One could take any Lorentz manifold, compute $(G + \Lambda \mathrm{id}\,)/(8\pi)$ and call the result the stress energy of the matter in that universe. This is not the intended use of the equations. Rather one should have an opinion what kind of matter is in the universe one intends to model, one should understand this matter well enough to be able to write down its stress energy tensor and finally look for a Lorentz manifold such that the Einstein Equation is satisfied. For how much complication should we be prepared? First, of course, there are the stars. It turned out that for modeling ordinary stars one does not need General Relativity. And the more exotic stars, imploding ones for example, require so broad a background in physics that they are out of my reach. We have seen the Schwarzschild geometry and glimpses of Kerr as models of the outside of a star. The next larger structures are galaxies and eventually the cosmology. I want to recall a very successful continuous model of an obviously discrete situation: *the kinetic theory of gases* in terms of differentiable functions called volume, pressure and temperature. A gas consists of molecules of diameter $10^{-10}m$ and up, and their mean distance is about a factor 30 larger. Our galaxy has a diameter of about 50.000 light years and the distance to the Andromeda galaxy is about 20 times that large. It will turn out that a cosmological model in which the matter is a dust of mass density $\rho$ and the dust grains are the galaxies (in other words: a very oversimplifying assumption) is surprisingly successful. And for the galaxies themselves, the ratio of distances between stars to star diameters is more like $10^7$ and therefore maybe too large for a continuous approximation. (I have been told that the shuttle reentry computations in the very thin high atmosphere do not describe the "gas" by using a very small continuous density, but really deal with individual molecules.) Very recently I obtained the following reference:

*1995 Phys. Rev. Letters 75, 3046 , Neugebauer, G.; Meinel, R.: General Relativistic Gravitational Field of a Rigidly Rotating Disk of Dust: Solution in Terms of Ultraelliptic Functions .*

I did not have time to see what one can learn from it, the words *"rigidly rotating"* do exclude that it is a galactic model. Concerning galaxies. I know that really huge numerical simulations have been made, but I do not know any details. Therefore, with obvious regret, I cannot discuss relativistic models of galaxies in these notes.

The remaining goal therefore is to discuss a family of cosmological models that are filled with a very simple type of matter. We will not meet complicated stress energy tensors, but in the same way as the detailed discussion of our first vacuum solution (Schwarzschild) turned out to be very educational we will gain insight about the interplay between matter and geometry on a cosmological scale even though we work with the simplest kind of matter that can be imagined. Before turning to that goal I end this section with definitions and with some more local arguments.

A matter is called a **perfect fluid** if it has just two physical properties called *pressure* $p$ and *mass density* $\rho$ ($p, \rho$ are differentiable functions) and if at every point in the rest system of the matter (this makes sense only where $\rho \neq 0$) the stress energy 1-1-tensor $T$ has the rest space as 3-dim eigenspace with eigenvalue $p$ and the time like unit vector $U$ of the rest frame is an eigenvector with eigenvalue $-\rho$. Since $U$ is defined everywhere, it is a time like unit vector field whose integral curves are the world lines of the matter particles. Note that the infinitesimal rest spaces $U^\perp$ in general are **not** an integrable distribution. This means that in general there are no natural space slices. This phenomenon will be obscured by our examples: additional simplicity assumptions make $U^\perp$ integrable and therefore lead to natural space slices. I find it important to emphasize that even with all the specifics above we do not yet have some physically specific perfect fluid. In addition one needs a

*matter equation* or *equation of state*: $F(p, \rho) = const,\ \frac{\partial}{\partial p} F \neq 0.$

We shall mainly work with the equation $p = 0$ that specifies a **dust**.
We shall mention $3p - \rho = 0$ specifying a perfect fluid called **photon gas**.
In the absence of a matter equation the following inequalities are required: $0 \leq 3p \leq \rho$.
My knowledge of continuum mechanics is insufficient for comments about these inequalities.

Next we translate the given information about $T$, using the Einstein equation, in information about $Ric$:

For arbitrary vectors $W$ holds: $\quad T \cdot W = (p \cdot W + (\rho + p) g(U, W) \cdot U$

$$8\pi \cdot \mathrm{trace}\,(T) = \mathrm{trace}\,(Ric) - 2\mathrm{trace}\,(Ric) - 4\Lambda$$

$$Ric = 8\pi (T - \frac{1}{2}\mathrm{trace}\,(T)\mathrm{id}\,) + \Lambda \mathrm{id}$$

$$Ric(U) = (\Lambda - 4\pi(\rho + 3p)) \cdot U, \quad Ric\big|_{U^\perp} = (\Lambda + 4\pi(\rho - p)) \cdot \mathrm{id}\,\big|_{U^\perp}.$$

By looking at the Ricci tensor we can now recognize whether some Lorentz manifold has as its matter content a perfect fluid. The quadratic examples of lecture 2 do not model such type of matter.

Recall that, when Einstein wrote down the above field equation, physicists had already met stress energy tensors of materials and they were convinced that $T$ would be divergence free for all materials. Therefore Einstein constructed the right side of the equation to be divergence free. We learn some facts about perfect fluids by computing the divergence of $T$:

$$\mathrm{div}\,(T) := \sum_i \frac{(D_{e_i} T) \cdot e_i}{g(e_i,\ e_i)} \Longrightarrow g(\mathrm{div}\,(T), W) = \sum_i \frac{g((D_{e_i} T) \cdot W, e_i)}{g(e_i,\ e_i)}$$

$$g(\mathrm{div}\,(T), W) = T_W p + (p + \rho) g(W, D_U U) + g(W, U) \mathrm{div}\,((p + \rho) U).$$

If we *use* div $(T) = 0$ and apply this computation for $W \perp U$, then we get

$$D_U U = -(\text{grad}\, p)/(p+\rho), \quad \text{grad } = \text{grad}^{\,Restspace}$$

in particular, in the case of dust, we get *geodesic world lines* for the dust particles. In general the acceleration is caused by the pressure gradient (in the rest space).
If we use the computation for $W = U$ in the dust case, we get div $(\rho \cdot U) = 0$, a conservation of mass result. This shows that quite basic facts about the behavior of the perfect fluid follow from the Einstein field equation without prior knowledge of these facts from classical physics.

### What is div $T = 0$ good for?

If in some field theory a vector field $V$ with div $(V) = 0$ occurs then Gauß' theorem implies that the flow of $V$ carries some conserved quantity around. However, there is no Gauß' theorem for 1-1-tensors and therefore: why is div $T = 0$ important? A celebrated fact from classical mechanics is the observation that symmetry groups, or Killing fields, lead to conserved quantities. And Killing fields $X$ (characterized by the skew-symmetry of their covariant differential, $DX = -DX^{tr}$) are similarly useful in our context:

Claim: div $T = 0$ and $DX = -DX^{tr} \implies V := T \cdot X$ satisfies div $(V) = 0$.

Proof:

$$DV = (DT) \cdot X + T \cdot DX,$$

$$\text{div}\,(V) = \text{trace}\,(DV),$$

$$\text{trace}\,(T \cdot DX) = 0 \ \text{ since } T \text{ is symmetric and } DX \text{ is skew,}$$

$$\text{trace}\,((DT) \cdot X) = \sum_i \frac{g((D_{e_i}T) \cdot X, e_i)}{g(e_i, e_i)} = \sum_i \frac{g((D_{e_i}T) \cdot e_i, X)}{g(e_i, e_i)} = g(\text{div}\,(T), X) = 0.$$

This shows that the divergence free stress energy tensor $T$ together with any Killing field $X$ leads to a divergence free vector fields $V = T \cdot X$, i.e. to vector fields $V$ whose flow transports some conserved quantity. This observation makes div $T = 0$ important, if there are Killing fields. Not surprisingly do our simplified models carry Killing fields, but on a real cosmology with all its individual features there won't be Killing fields. Is div $T = 0$ still important? I will argue "yes, and for almost the same reason".

First recall that in Euclidean space and in Minkowsky's Special Relativity Killing fields are explicitly determined by value and derivative at one point:

$$X(x) = X(p) + DX\big|_p \cdot (x - p).$$

Secondly, an observing physicist, of course, cannot leave his world line. Moreover we have by now some experience in viewing physicists as infinitesimal observers who perform their experiments in the tangent spaces of the Lorentz manifold, along their world line. This means that for observing conserved quantities they do not really need globally defined Killing fields, what they need are "almost" Killing fields defined on a tube around their world line. Recall that a Killing field satisfies along any geodesic $\gamma$(i.e. along the world line of any unaccelerated observer) and for any *parallel* field $v$ along $\gamma$ the following PDE:

$$D_{\gamma'}(D_v X) + R(X, \gamma')v = 0.$$

This says: $X$ and $DX$ is determined along $\gamma$ by its initial value $X(\gamma(0))$ and its initial derivative $DX\big|_{\gamma(0)}$, just as in the Euclidean/Minkowski case. Of course $DX\big|_{\gamma(0)}$ needs to be skew-symmetric, but if this initial constraint is met then $DX\big|_{\gamma(s)}$ continues to be skew-symmetric:

$$\frac{d}{ds} g(D_{v(s)}X, v(s)) = -g\Big(R\big(X(s), \gamma'(s)\big)v(s),\ v(s)\Big) = 0.$$

We can therefore construct as many almost Killing fields $X$ on an infinitesimal tube around $\gamma$ as we have in Special Relativity and $\operatorname{div} T = 0$ allows us to observe the conserved quantities of the flows of the fields $V := T \cdot X$, so that $\operatorname{div} T = 0$ is really responsible for observable conserved quantities.

**Interplay with Conformal Flatness.**

We are interested in conformally flat Lorentz manifolds because then we get solutions of Maxwell's equation for free. A (pseudo)-Riemannian metric is (locally) conformally flat iff its Weil conformal curvature tensor vanishes. In such a case one can write the full curvature tensor in terms of the Ricci tensor. In the case of a perfect fluid we saw that the Ricci tensor does not distinguish any space like directions in the rest spaces of the matter. Taking the two facts together shows:

*A conformally flat perfect fluid is curvature isotropic.*

We write more explicitly what we mean by "curvature isotropic with respect to $U$", i.e., by the property that *the curvature tensor distinguishes no directions in the rest spaces $U^\perp$ of the matter.* Clearly, such a curvature tensor has to have the following properties:

$$\begin{aligned} X, Y, Z \perp U \Longrightarrow R(X, Y)Z &= k(p)(g(Y, Z)X - g(X, Z)Y), \\ R(X, U)U &= \mu(p) \cdot X, \end{aligned}$$

with the immediate consequences:

$$\begin{aligned} R(X, Y)U &= 0, \\ R(U, X)Y &= -\mu(p) \cdot g(X, Y) \cdot U. \end{aligned}$$

(Note that $g(R(U, X)Y, Z) = 0$ for all $Z \perp U$ and $g(R(U, X)Y, U) = g(R(X, U)U,\ Y)$.)
This is enough information about the curvature tensor to check that any curvature isotropic curvature tensor has its Weyl conformal curvature tensor vanish, so that the manifold is locally conformally flat. Moreover, we find for the Ricci tensor (of such a curvature tensor):

$$\begin{aligned} ric(U, U) &= 3\mu(p) &&= -\lambda_U = (-\Lambda + 4\pi(\rho + 3p)) \\ ric(U, Y) &= 0 \\ ric(X, Y) &= (2k - \mu)g(X, Y) &&= \lambda_{U^\perp} = (\Lambda + 4\pi(\rho - p)). \end{aligned}$$

This shows that the eigenspace decomposition is the correct one for a perfect fluid (we also need to satisfy $0 \le 3p \le \rho$), so that, essentially, "conformally flat perfect fluid" and "curvature isotropic space" describe the same Lorentz manifolds.
Note:

$$6k - 2\Lambda = 16\pi\rho, \quad 4\mu - 2k + 2\Lambda = 16\pi p, \quad \mu + k = 4\pi(p + \rho).$$

After introducing the concepts and show immediate relations we come to a real theorem:

**Theorem of Schur type.** Let $M^4$ be curvature isotropic for a time like unit vector field $U$ so that $M^4$ models a perfect fluid. We also assume $\rho > 0$, since otherwise one cannot everywhere define the local rest frame of the matter, namely $U, U^\perp$. Then:
a) $U^\perp$ is an **integrable** distribution.
b) The 3-dim integral manifolds have intrinsically constant curvature.
c) A matter equation $F(p, \rho) = 0$, $\frac{\partial}{\partial p} F \neq 0$ implies $D_U U = 0$ so that extrinsically these integral manifolds are parallel hypersurfaces with the matter world lines as the orthogonal geodesics.

The **proof** is modeled after Schur's theorem for Riemannian manifolds that states: *If the sectional curvatures are constant at each point then they are constant.* The argument relies on the 2nd Bianchi identity, we will use

$$0 = (D_U R)(X, Y)Z + (D_X R)(Y, U)Z + (D_Y R)(U, X)Z.$$

(Other combinations of arguments do not contain additional information.) Our curvature assumptions are such that the orthogonal splitting $T_pM = U(p)\mathbb{R} \oplus U^\perp$ is essential. Therefore we will use the induced covariant derivative $D^\perp$ on the 3-dim bundle $U^\perp$ over $M$. By $X, Y, Z$ we will always denote vector fields from that bundle.

$$D^\perp X := DX + g(DX, U) \cdot U \perp U.$$
$$D^\perp_{\dot c} X = 0 \;\Rightarrow\; D_{\dot c} X = -g(D_{\dot c} X, U) \cdot U = g(X, D_{\dot c} U) \cdot U.$$

Clearly, $D^\perp$-parallel vector fields have constant scalar products. For the evaluation of the terms in the Bianchi sum we may assume that the vector fields $X, Y, Z \perp U$ are $D^\perp$-parallel in the direction of the differentiation field. Now compute the Bianchi sum terms:

First: $D_U(R(X, Y)Z) = dk(U)(g(Y, Z)X - g(X, Z)Y) + k(g(Y, Z)D_U X - g(X, Z)D_U Y)$.

Since $D_U X, D_U Y, D_U Z$ are proportional to $U$ we have

$$R(X, Y)D_U Z = 0, \quad R(D_U X, Y)Z = -\mu g(Y, Z)D_U X. \quad R(X, D_U Y)Z = \mu g(X, Z)D_U Y$$
$$(1) \quad (D_U R)(X, Y)Z = dk(U)(g(Y, Z)X - g(X, Z)Y) \quad \perp U$$
$$+ (k + \mu)(g(Y, Z)D_U X - g(X, Z)D_U Y) \quad \in U\mathbb{R}$$

Second: $$D_X(R(U, Y)Z) = -d\mu(X)g(Y, Z) \cdot U - \mu g(Y, Z)D_X U.$$

Again, the derivatives of the arguments are either parallel or orthogonal to $U$, hence

$$R(D_XU,Y)Z = k(g(Y,Z)D_XU - g(D_XU,Z)Y), \quad R(U,D_XY)Z = 0,$$
$$R(U,Y)D_XZ = -\mu g(Z,D_XU)Y \quad \text{(recall } D_XZ = g(Z,D_XU)U\text{ )}$$
$$(2)\quad (D_XR)(Y,U)Z = -(D_XR)(U,Y)Z$$
$$= d\mu(X)g(Y,Z)U \quad \in U\mathbb{R}$$
$$+ (k+\mu)(g(Y,Z)D_XU - g(Z,D_XU)Y) \quad \perp U.$$

And similarly (interchange $X$ and $Y$ and a sign)

$$(3)\quad (D_YR)(U,X)Z = -d\mu(Y)g(X,Z)U \quad \in U\mathbb{R}$$
$$- (k+\mu)(g(X,Z)D_YU - g(Z,D_YU)X) \quad \perp U.$$

Using the 2nd Bianchi identity in (1)+(2)+(3) gives two equations, one in $U\mathbb{R}$, one in $U^\perp$:

In $U\mathbb{R}$ $\;d\mu(X)g(Y,Z)U - d\mu(Y)g(X,Z)U = -(k+\mu)(g(Y,Z)D_UX - g(X,Z)D_UY)$,

In $U^\perp$ $\;dk(U)(g(Y,Z)X - g(X,Z)Y) =$

$$= (k+\mu)\left(g(Y,Z)D_XU - g(X,Z)D_YU + g(D_YU,Z)X - g(D_XU,Z)Y\right).$$

If we use unit vectors $X \perp Y = Z$ in the first equation we get

$$d\mu(X) = -(k+\mu)g(X,D_UU),$$

we computed earlier

$$\operatorname{div}(T) = 0 \quad \Longrightarrow \quad 8\pi dp(X) = -8\pi(\rho+p)g(X,D_UU) =$$
$$= d(2\mu - k)(X) = -2(\mu+k)g(X,D_UU),$$

and both equations together give

$$dk(X) = 0 \quad \text{for all} \quad X \in U^\perp$$

| |
|---|
| This shows: if $\rho$, hence $k$, are not konstant then the levels of $\rho$ are the integral manifolds of the distribution $U^\perp$. |

We still have to consider the case of constant $k$ since the absence of matter equations makes still many examples possible. Therefore we need another proof of the integrability of the distribution $U^\perp$. We claim, the vector field $(k+\mu)U$ has a symmetric covariant differential and therefore is (locally) the gradient of a function, and since $(k+\mu) > 0$ is implied by our assumption $\rho > 0$, this proves integrability of $U^\perp$. To see the claim, first put orthonormal vectors $X, Y, Z$ in the second part of the above Bianchi equation to obtain

$$0 = (k+\mu)(g(D_YU,Z)X - g(D_XU,Z)Y), \quad \text{hence} \quad g(D_YU,Z) = 0.$$

This says that for any orthonormal basis in $U^\perp$ the matrix of $DU\big|_{U^\perp}$ is diagonal, in particular symmetric. It remains to check, with the above equations, the remaining symmetry:

$$g(D_X((k+\mu)U),U) = g(d\mu(X)U,U) = (k+\mu)g(X,D_UU) = g(D_U((k+\mu)U),X),$$

and thus prove the integrability of $U^\perp$ in all cases. We emphasize that this integrability was deduced from strong assumptions, it is normally false.

Next we determine the *intrinsic curvature* of the integral submanifolds of $U^\perp$. We also refer to them as *space slices.* The unit (time like) vector field $U$ is of course normal along them. The Weingarten map (=shape operator) of the space slices therefore is $S := DU$ and we proved already that $DU$ is diagonal in any orthonormal basis, i.e., is proportional to id. We use unit vectors $X \perp Y = Z$ in the Bianchi equation involving $dk(U)$. Taking a scalar product with $X$ we obtain:

$$-\frac{dk(U)}{k+\mu} \cdot g(Y,Y) = g(D_X U, X) + g(D_Y U, Y) = 2 \cdot \text{eigenvalue of } S.$$

We use this in the Gauss equation:

$$R(X,Y)Z = k(g(Y,Z)X - g(X,Z)Y) \quad (\text{ assumption about } M^4)$$
$$\overset{(Gauss)}{=} R^{Hyp}(X,Y)Z - ((g(SY,Z)SX - g(SX,Z)SY) \cdot g(U,U)^{-1}$$
$$R^{Hyp}(X,Y)Z = \left(k - \frac{1}{4}(\frac{dk(U)}{k+\mu})^2\right) \cdot (g(Y,Z)X - g(X,Z)Y).$$

This shows that the space slices satisfy the assumptions of the Riemannian Schur theorem so that the *curvature value is indeed constant* on each space slice.

Finally we assume a matter equation $F(\rho,p) = 0$, $\frac{\partial}{\partial p}F \neq 0$. Recall that we proved for all $X \perp U$ that $dk(X) = 0$. This says that $\operatorname{grad}\rho$ is proportional to $U$ (including 0). Differentiation of the matter equation gives that $\operatorname{grad} p$ is proportional to $\operatorname{grad}\rho$ (again including 0). Therefore we have for all $X \perp U$ that $0 = dp(X)$, hence

$$0 = dp(X) = -(\rho + p)g(X, D_U U).$$

The integral curves of $U$, the world lines of the matter particles, are therefore *geodesics* with integrable orthogonal complements $U^\perp$ and these space slices are a family of *geodesically parallel hypersurfaces.* Q.E.D.

## Summary of Conformal Changes

Given $\bar{g} = \lambda^{-2} g$

then $\overline{D}_Y Z = D_Y Z + \Gamma(Y,Z)$

with $\Gamma(Y,Z) = -\frac{T_Z\lambda}{\lambda} Y - \frac{T_Y\lambda}{\lambda} Z + g(Y,Z)\mathrm{grad}\,\lambda.$

$\overline{R}(X,Y)Z = R(X,Y)Z + (D_X\Gamma)(Y,Z) - (D_Y\Gamma)(X,Z) + \Gamma(X,\Gamma(Y,Z)) - \Gamma(Y,\Gamma(X,Z))$

gives with the abbreviation $B := D\mathrm{grad}\,\lambda$

$$\begin{aligned}\overline{R}(X,Y)Z = R(X,Y)Z &+ \frac{1}{\lambda}\big(g(Y,Z)BX - g(X,Z)BY + (g(BY,Z)X - g(BX,Z)Y\big)\\ &- \frac{1}{\lambda^2} g(\mathrm{grad}\,\lambda, \mathrm{grad}\,\lambda) \cdot (g(Y,Z)X - g(X,Z)Y).\end{aligned}$$

This gives the new Einstein tensor as

$$\overline{G} = \lambda^2 \Big( G + \frac{2}{\lambda} B + \big(\frac{3}{\lambda^2} g(\mathrm{grad}\,\lambda, \mathrm{grad}\,\lambda) - \frac{2\Delta\lambda}{\lambda}\big) \cdot \mathrm{id} \Big).$$

We have done no computations with the *Weyl conformal curvature tensor*, we list it as a reference:

$$\begin{aligned}&\overline{C}(X,Y)Z = C(X,Y)Z =\\ &= R(X,Y)Z - \frac{1}{n-2}\big(ric(Y,Z)X - ric(X,Z)Y + g(Y,Z)Ric(X) - g(X,Z)Ric(Y)\big)\\ &\qquad + \frac{\mathrm{trace}\,(Ric)}{(n-2)(n-1)}\big(g(Y,Z)X - g(X,Z)Y\big).\end{aligned}$$

## Cosmological Models

Infinitesimally isotropic dust, red shift - luminosity, mass per red shift

We will not derive the standard cosmological model assuming Schur's 2nd Bianchi arguments from the previous lecture. But we will derive this family of models from scratch, with simpler arguments, following the historic route. This approach needs stronger assumptions, fortunately assumptions that are the conclusions of the Schur theorem. I also prefer to skip the factor $8\pi$ from the Einstein equation, i.e. the functions $\rho, p$ in this section are the functions $8\pi\rho, 8\pi p$ of the last section.

**Model assumptions, for Friedman or Robertson-Walker universes:**

Matter content.

The matter of the model is a perfect fluid. Mostly we assume the matter equation for **dust**, $p = 0$. To illustrate how the type of matter changes the model we will also deal with the matter equation for a **photon gas**, $3p = \rho$.

Symmetry.

Other observers on matter world lines should see the universe as we do, and, roughly speaking, the observations do not distinguish special rest space directions (i.e. orthogonal to matter world lines). We turn this into the assumption: *the curvature tensor distinguishes no directions in the rest spaces.*

Ansatz.

From these assumptions we concluded that the matter world lines are geodesics and that the orthogonal distribution is integrable, giving space slices of constant intrinsic and extrinsic curvature. This foliation also defines a global time function $\tau$ and the curvatures as well as $\rho$ and $p$ depend on $\tau$.

The underlying manifold therefore is

$$M^4 = M^3_\kappa \times (a, b),$$

with $(a, b)$ to be determined and $M^3_\kappa$ a space of constant curvature $\kappa$.

$M^4$ has a warped product metric

$$\bar{g} = a^2(\tau) g_\kappa(.,.) - d\tau^2.$$

We assume $a(\tau = \mathit{today}) = 1$ so that $g_\kappa$ is the metric of the space slice that cuts our world line at $\tau = \mathit{today}$. Note that I take $\kappa$ as a continuous curvature parameter and not, as in part of the literature, $\kappa = -1, 0, +1$ only.

*What do the Einstein equations say about the scaling function $a(\tau)$?*

Let $U$ be the timelike unit tangent field to the matter world lines ($\tau$-lines) and let $X, Y, Z \perp U$ be tangential to the space slices. If $X$ is parallel along the $\tau$-lines then $J(\tau) := a(\tau)X$ is a Jacobi field. This gives

1. $\overline{R}(X, U)U = -\dfrac{a''}{a} X$ (Jacobi equation for $aX$),

2. $\overline{S}(X) = \dfrac{a'}{a} X$ (Shape operator of space slice),

3. $$\overline{R}(X,Y)Z = \frac{1}{a^2}R_\kappa(X,Y)Z - \big(\bar g(\overline{S}Y,Z)\overline{S}X - \bar g(\overline{S}X,Z)\overline{S}Y\big)\bar g(U,U)^{-1}$$
$$= (\frac{\kappa}{a^2} + \frac{a'^2}{a^2})\big(\bar g(Y,Z)X - \bar g(X,Z)Y\big), \quad \text{(Gauss eq.)}$$

4. $$\overline{Ric}(U) = \frac{3a''}{a}U \qquad \overline{Ric}(Z) = (\frac{2\kappa + 2a'^2}{a^2} + \frac{a''}{a})Z$$

5. $$\frac{1}{2}\text{trace}\,(\overline{Ric}) = \frac{3a''}{a} + 3\frac{\kappa + a'^2}{a^2}$$

Einstein eq.: $$(\overline{G} + \Lambda)\cdot U = (-3(\frac{\kappa + a'^2}{a^2}) + \Lambda)U = -\rho U$$
$$(\overline{G} + \Lambda)\cdot Z = (-\frac{\kappa + a'^2}{a^2} - \frac{2a''}{a} + \Lambda)Z = pZ$$

Dust: $$\rho = 3(\frac{\kappa + a'^2}{a^2}) - \Lambda \qquad -\frac{\kappa + a'^2}{a^2} - \frac{2a''}{a} + \Lambda = 0$$

We simplify the system to a first order ODE by observing a first integral:

$$\frac{1}{3}\big(\rho(\tau)a(\tau)^3\big)' = \big(aa'^2 + \kappa a - \frac{\Lambda}{3}a^3\big)' = a^2a'\left(\frac{2a''}{a} + \frac{\kappa + a'^2}{a^2} - \Lambda\right) = 0$$
$$\rho(\tau)\cdot a(\tau)^3 = const. = \rho(today)\cdot 1, \quad \text{ODE: } a'^2 = \frac{\rho(today)}{a} - \kappa + \frac{\Lambda}{3}a^2.$$

This relation between the mass density and the scaling size agrees with our 3-dimensional intuition. It is also a good result because the mass density is more directly observable than our Ansatz function $a(\tau)$, the scaling size of the space slices.

---

We pause briefly for a comparison with the computations in the Schur theorem. There the curvature function $k$ of the 4-dimensional curvature tensor was used. Equation 3 above expresses $k$ in terms of $\kappa$ and $a$, the relation $2(k+\mu) = (\rho + p)$ is from the previous section and the Einstein equations for $M^4$ (above) give $(\rho + p)$ in terms of $\kappa$ and $a$:

$$k \overset{(3.)}{=} \frac{\kappa}{a^2} + \frac{a'^2}{a^2} \quad \text{and} \quad 2(k+\mu) = (p+\rho) \overset{(Einstein)}{=} 2\frac{\kappa + a'^2}{a^2} - \frac{2a''}{a}$$
$$\frac{1}{4}\left(\frac{dk(U)}{k+\mu}\right)^2 \overset{(2.)}{=} (\frac{a'}{a})^2 \overset{(Gauss)}{=} k - \frac{\kappa}{a^2} = curv(M^4) - curv(spaceslice^3).$$

One sees that the final result of the Schur argument agrees with the present more direct application of the Einstein equations.

---

We know from the previous lecture that the models under consideration are conformally flat. It is easier to deal with red shift predictions and application of Maxwell's equation in a conformally flat description of the model. It will also turn out that the resulting differential equations can be integrated one step more in the conformal description than in the above first approach. Therefore we will start again from the beginning, but for further comparison with the physics literature we will also come back to this first approach.

We introduce a new time function $t$ and define what will turn out to be the conformal factor:

$$dt := \frac{d\tau}{a(\tau)} \quad \text{and} \quad \lambda(t) := a(\tau(t))^{-1}. \quad d\tau = \frac{dt}{\lambda(t)}. \qquad \text{Note } \lambda(today) = 1.$$

This transforms the above Ansatz metric in a conformally flat form:

$$\bar{g} = a(\tau)^2 g_\kappa - d\tau^2 = \lambda(t)^{-2}(g_\kappa - dt^2).$$

From the definition of $t, \lambda(t)$ follows (with $\frac{d}{d\tau}h(\tau) = h'(\tau), \ \ \frac{d}{dt}h(t) = \dot{h}(t)$):

$$\frac{a'}{a}(\tau) = -\dot{\lambda}(t), \qquad \frac{a''}{a} = -\lambda\ddot{\lambda} + \dot{\lambda}^2.$$

These relations suffice to translate the (Einstein) differential equations for $a(\tau)$ into differential equations for $\lambda(t)$. We will use this only as a check and derive the equations for $\lambda(t)$ from scratch, using this as another illustration how the Einstein equations lead in the presence of matter equations to a model description in terms of explicit differential equations.

Curvature tensor, Ricci tensor and Einstein tensor for the product metric $g = g_\kappa - dt^2$ are easily obtained (observe that $U$ is globally parallel for $g$):

$$\begin{aligned}
&R(*,*)U = 0, \quad R(X,Y)Z = \kappa(g(Y,Z)X - g(X,Z)Y),\\
&Ric(U) = 0, \quad Ric(X) = 2\kappa X, \quad \frac{1}{2}\text{trace}\,(Ric) = 3\kappa,\\
&G(U) = -3\kappa U, \quad G(X) = -\kappa X.
\end{aligned}$$

For the conformally changed metric $\bar{g} = \lambda^{-2} g$ we compute the Einstein tensor with the conformal-change-formula at the end of last lecture.
Note $\text{grad}_g\lambda = -\dot{\lambda}U, \ \ D\text{grad}_g\lambda = -\ddot{\lambda}g(U,.)U$

$$(\overline{G} + \Lambda)(X) = \big(\lambda^2(-\kappa + 0 - \frac{3\dot{\lambda}^2}{\lambda^2} + \frac{2\ddot{\lambda}}{\lambda} + \Lambda\big)X \overset{(!)}{=} pX$$

$$(\overline{G} + \Lambda)(U) = \big(\lambda^2(-3\kappa - \frac{2\ddot{\lambda}}{\lambda} - \frac{3\dot{\lambda}^2}{\lambda^2} + \frac{2\ddot{\lambda}}{\lambda} + \Lambda\big)U \overset{(!)}{=} -\rho U$$

This gives the expected differential equations (compare those for $a(\tau)$):

$$\begin{aligned}
&2\lambda\ddot{\lambda} - 3\dot{\lambda}^2 - \kappa\lambda^2 + \Lambda = 0,\\
&\rho(t) = 3\dot{\lambda}^2 + 3\kappa\lambda^2 - \Lambda = 2\kappa\lambda^2 + 2\lambda\ddot{\lambda}.
\end{aligned}$$

Hence: $\quad \dot{\rho} = 6\dot{\lambda}(\kappa\lambda + \ddot{\lambda}) = 3\frac{\dot{\lambda}}{\lambda}\rho,$

and: $\quad \rho(t) = \rho(T)\cdot\lambda(t)^3, \quad$ Abbreviate $T := today$ henceforth.

As in the first description $\rho(t)$ scales expectedly with $\lambda(t)^3$ so that scaling sizes of space slices that intersect matter world lines at $\gamma(t)$ can equivalently be expressed in terms of matter densities, more precisely $\rho(t)^{1/3}$, along $\gamma$.

So far the two descriptions show the same level of complication, the advantages of the conformal description begin now. The just established fact that $\rho(t)\lambda(t)^{-3}$ is a constant translates into a first order ODE for $\lambda$ that has the two Einstein equations we started with as consequences (recall that this statement holds also in the $a(\tau)$-description):

$$\frac{d}{dt}(\frac{1}{3}\rho(t)\lambda(t)^{-3}) = \frac{d}{dt}\big(\dot\lambda^2\lambda^{-3} + \kappa\lambda^{-1} - \frac{\Lambda}{3}\lambda^{-3}\big) = 2\dot\lambda\ddot\lambda\lambda^{-3} - 3\dot\lambda^3\lambda^{-4} - \kappa\dot\lambda\lambda^{-2} + \Lambda\dot\lambda\lambda^{-4}$$
$$= \dot\lambda\lambda^{-4}\big(2\lambda\ddot\lambda - 3\dot\lambda^2 - \kappa\lambda^2 + \Lambda\big) = 0.$$

So finally we have reached

**The Equation of the Cosmological Model**

$$\dot\lambda^2 = \frac{\rho(T)}{3}\cdot\lambda^3 - \kappa\lambda^2 + \frac{\Lambda}{3}, \quad \rho(t) = \rho(T)\cdot\lambda(t)^3,$$
$$\overline{g} = \frac{1}{\lambda^2}(g_\kappa - dt^2) = \left(\frac{\rho(T)}{\rho(t)}\right)^{2/3}\cdot(g_\kappa - dt^2).$$

For $\Lambda \neq 0$ this ODE for $\lambda(t)$ is the ODE of an elliptic function while for $\Lambda = 0$ an explicit integration in terms of elementary transcendental functions is possible. We therefore assume in the following $\Lambda = 0$ whenever reference to the explicit solution is made. We use abbreviations for $\sin(\sqrt{\kappa}\,t)/\sqrt{\kappa}$ and similar functions as follows:

$$\mathbf{s}_\kappa'' + \kappa\mathbf{s}_\kappa = 0, \quad \mathbf{s}_\kappa(0) = 0, \quad \mathbf{s}_\kappa'(0) = 1. \qquad \text{Note: } (\mathbf{s}_\kappa')^2 + \kappa\mathbf{s}_\kappa^2 = 1$$
$$\mathbf{c}_\kappa'' + \kappa\mathbf{c}_\kappa = 0, \quad \mathbf{c}_\kappa(0) = 1, \quad \mathbf{c}_\kappa'(0) = 0. \qquad (\mathbf{c}_\kappa')^2 + \kappa\mathbf{c}_\kappa^2 = \kappa$$
$$\mathbf{c}_\kappa = \mathbf{s}_\kappa'$$

**Claim.** In the case $\Lambda = 0$ we have the following **explicit solution** of the model ODE:

$$\lambda(t) := \mathbf{s}_\kappa(\frac{T-t_0}{2})^2\cdot\mathbf{s}_\kappa(\frac{t-t_0}{2})^{-2} \quad \text{(Recall } T = \textit{today}\text{)}.$$

With $$\rho(T) = 3\mathbf{s}_\kappa(\frac{T-t_0}{2})^{-2}, \quad \rho(t) = 3\mathbf{s}_\kappa(\frac{T-t_0}{2})^4\cdot\mathbf{s}_\kappa(\frac{t-t_0}{2})^{-6}.$$

Here $t_0$ is the time where the mass density becomes infinite. There is no harm in setting $t_0 = 0$. To prove the claim compute $(\dot\lambda)^2/\lambda^2 + \kappa$ with the help of $(\mathbf{s}_\kappa')^2 + \kappa\mathbf{s}_\kappa^2 = 1$ and find it equal to $(\mathbf{s}_\kappa(T/2)^{-2}\cdot\lambda(t)$, hence $\rho(T)/3 = \mathbf{s}_\kappa(T/2)^{-2}$.

As a first observation we have a **Big Bang prediction**: If we go backwards in time and reach infinite mass density at $t = t_0 = 0$ then this moment is the Big Bang for the forward time development of the model. - This statement requires one word of caution: Before the mass density reaches infinity it becomes so large that the matter presumably can no longer be treated as a dust. In other words, the model assumptions become invalid before the Big Bang is reached. At the time when the dust assumption becomes invalid one may change the matter equation to that of a photon gas and compute somewhat further back in time until these matter equations become invalid. It is generally believed that one can keep adjusting the matter equations as one gets arbitrarily close to the Big Bang.

Clearly, the Big Bang prediction has caused a lot of excitement although the Big Bang is, strictly speaking, out of observational reach. We come now to a **red shift prediction** which is also exciting because it concerns one of the most dominant observational facts from astronomy.

For the physically unimportant product metric $g = g_\kappa - dt^2$ we have that the vector field $U$ is a time like Killing field of constant length. Therefore we have no red shift between observers represented by the integral curves of $U$. Under the conformal change to the physically relevant metric $\bar{g} = \frac{1}{\lambda^2} g$ these integral curves become the world lines of the matter particles of that model. We have computed the red shift caused by a conformal change and found:

$$1 + z = \frac{\omega_{Source}}{\omega_{Observer}} = \frac{\lambda(t)}{\lambda(T)} = \frac{a(\tau = today)}{a(\tau(t))} = \frac{\mathbf{s}_\kappa(T/2)^2}{\mathbf{s}_\kappa(t/2)^2} = \left(\frac{\rho(t)}{\rho(T)}\right)^{1/3}.$$

This has an immediate interpretation: The red shift of light received from 'distant' galaxies tells us how much denser the universe was at time $t$ of emission than at time $T = today$ of reception. (In particular, the red shift from the Big Bang is infinite, which certainly adds to the observational difficulties.)
The historical and much more common interpretation is different:
if one interprets $\tau(today) - \tau(t) = \Delta\tau$ as the travel time of the light from the space slice at emission to us then this difference can also be called "distance" between the emitting star and us. The first order Taylor formula gives (recall $a(today) = 1$)

$$z \approx \frac{\Delta a}{a} \approx \frac{a'}{a}\Delta\tau.$$

This says that the red shift increases linearly with the "distance". Finally, the simplest interpretation of red shift is the Doppler shift caused by relative motion. In other words: the relative velocity between us and the galaxies increases with their distance! This is the *expansion of the universe* observation.

Whichever interpretation of the red shift prediction one prefers: the model has clearly made contact with observational facts.

---

We interrupt the discussion of the model to look at another matter equation, at a **photon gas,** $\rho = 3p$. If we insert this into the above eigenvalue computations for the Einstein tensor in the conformally flat description we get

(1) $$\frac{\rho(t)}{3} = 2\lambda\ddot{\lambda} - 3\dot{\lambda}^2 - \kappa\lambda^2 + \Lambda,$$

(2) $$\rho(t) = 3\dot{\lambda}^2 + 3\kappa\lambda^2 - \Lambda,$$

(1) + (2) $$\frac{4}{3}\rho = 2\lambda\ddot{\lambda} + 2\kappa\lambda^2.$$

$((1) - (2)/3)\dot{\lambda}/\lambda^5$ $$0 = \frac{d}{dt}\left(\frac{\dot{\lambda}^2}{\lambda^4} + \frac{\kappa}{\lambda^2} - \frac{\Lambda}{3\lambda^4}\right) = \frac{d}{dt}\frac{\rho(t)}{3\lambda(t)^4},$$

Differentiating (2) $\qquad \dot\rho = 6\dot\lambda(\kappa\lambda + \ddot\lambda) = 4\frac{\dot\lambda}{\lambda}\rho.$

Finally: $\qquad \rho(t) = \rho(T)\cdot\lambda(t)^{\mathbf{4}}.$

Again we end up with a first order ODE for the scaling function $\lambda(t)$, but a *different* power dependence, $\rho(t) \sim \lambda(t)^{\mathbf{4}}$, than for dust. Vice versa, this power law for $\rho$ and the first order ODE for $\lambda(t)$ imply the two Einstein equations.

---

For comparison with the literature we need to discuss the model parameters. One parameter is the cosmological constant $\Lambda$, but I do not know how to discuss its connection with observations. Our Ansatz had todays space slice curvature $\kappa$ as one model parameter, and the integration gave a second parameter, either the age $T$ of the universe or equivalently the matter density today, $\rho(T)$. None of these parameters is used in the literature. The expanding universe discussion suggests why the *Hubble function*$(\tau) = a'(\tau)/a(\tau) = -\dot\lambda(t)$ was defined. Its value today is the *Hubble Constant* $H$. It is one of the most prominent astronomical constants and it is one of the usual model parameters. We have:

$$H^2 := \dot\lambda(T)^2 \stackrel{(ODE)}{=} -\kappa + \frac{\rho(T)+\Lambda}{3}\ \Big|_{\Lambda=0} = \left(\frac{\dot{\mathbf{s}}_\kappa}{\mathbf{s}_\kappa}(T/2)\right)^2 = -\kappa + \frac{1}{\mathbf{s}_\kappa^2(T/2)}.$$

Clearly one can introduce $H$ instead of any of the other parameters to specify the model in the family. The second model parameter in the physics literature also comes from sympathy for Taylor approximations. The parameter is called *acceleration parameter* $q$ and defined via $a''|_{today}$ (recall $a''/a = -\lambda\ddot\lambda + \dot\lambda^2$ and $2\lambda\ddot\lambda = 3\dot\lambda^2 + \kappa\lambda^2 - \Lambda = \rho(t) - 2\kappa\lambda(t)^2$):

$$q := -\frac{a''}{a}\Big|_{today}\cdot\frac{1}{H^2} = \frac{\ddot\lambda}{\lambda}\left(\frac{\lambda}{\dot\lambda}\right)^2 - 1 = \frac{1}{2\dot\lambda^2}(\dot\lambda^2 + \kappa\lambda^2 - \Lambda) = \frac{1}{6H^2}(\rho(T) - 2\Lambda),$$
$$(2q-1)H^2 = \kappa - \Lambda.$$

With these equations one can choose, in terms of which parameters one wants the model to be specified. To me, $H$ and $\rho(T)$ seem closest to direct observations.

Some Comments

Of course one asks how model parameters could be determined experimentally, for example what is the sign of the curvature $\kappa$, how large is $\Lambda$? Note that the above two predictions, as impressive as they are, are qualitative, the predictions are the same for a large set of model parameters. One task therefore is, to find quantitative model predictions, see below.

One should not forget that the assumptions to derive this family of models were really strong: since we can see galaxies and empty space between them a mass density that is *constant* on space slices, is a rather drastic simplification. This is also true on a larger scale: clusters of galaxies and regions that are underpopulated are being observed. This is unfortunate since there is no relaxed version of Schur's theorem: *if the assumptions are almost satisfied then the conclusions are almost true.* Of course nobody expects from any theory to predict where clusters of galaxies would show up. But in the same way as earth' gravity and weather conditions might allow to say something about height, steepness and abundance of mountain ranges, one could hope to derive statistical properties of galaxy distribution – from much more complicated models than the presented family.

### Quantitative model predictions

The following discussion makes use of the explicit model obtained for $\Lambda = 0$. After one has seen what kind of predictions can be made, one can compute numerically similar predictions also for $\Lambda \neq 0$. The red shift prediction that we derived above

$$1 + z = \frac{\omega_{Source}}{\omega_{Observer}} = \frac{\lambda(t)}{\lambda(T)} = \frac{\mathbf{s}_\kappa(T/2)^2}{\mathbf{s}_\kappa(t/2)^2} = \left(\frac{\rho(t)}{\rho(T)}\right)^{1/3}.$$

is not a quantitative prediction as long as we do not know the time $t$ of emission. We want to combine it with a distance measurement to get a quantitative model prediction. We will compute a luminosity prediction for emission at time $t$ and eliminate $t$ from the two formulas to get a **red shift - luminosity prediction**. We begin the discussion with the Faraday form for a dipole field in Special Relativity, written in polar coordinates:

$$F_\infty = \cos(\omega(r - t)) \cdot \sin\vartheta \cdot (dt - dr) \wedge d\vartheta.$$

The first Maxwell equation is satisfied exactly, $d\,F_\infty = 0$, while the second Maxwell equation is only asymptotically satisfied since the above dipole term is only the leading term of the exact solution: $d{*}F_\infty \sim 1/r^2$, We started the description of the cosmological model with the conformally flat metric $g = g_\kappa - dt^2$, in polar coordinates:

$$ds^2 = dr^2 + \mathbf{s}_\kappa(r)^2(d\vartheta^2 + \sin^2\vartheta\ d\varphi^2) - dt^2.$$

For this metric we compute $d{*}F_\infty \sim 1/\mathbf{s}_\kappa^2(r)$. Therefore we interpret $F_\infty$ still as an asymptotic solution. (For $\kappa > 0$ there is no asymptotic behaviour as $r \to \infty$ and we should really use the exact solution.) In the same way as in Special Relativity the intensity of the outgoing radiation drops off as

$$(\text{area of distance sphere})^{-1} = \frac{1}{4\pi\mathbf{s}_\kappa(r)^2}.$$

Since the velocity of light is 1 we have $r = T - t$.

Next we have to make the conformal change to the physically relevant metric $\overline{g} = \lambda(t)^{-2}g$. We normalized the description so that $\lambda(T) = 1$ and therefore $\lambda(t) = 1 + z$. We have

$$g(e_j, e_j) = 1 \ \Rightarrow\ \overline{g}(\lambda e_j, \lambda e_j) = 1 \ \Rightarrow\ \bar{e}_j = \lambda e_j,$$
$$E_j = F(e_j, e_4),\ \overline{E}_j = F(\bar{e}_j, \bar{e}_4) \quad \Rightarrow \overline{E}_j = \lambda^2 E_j.$$

This allows to compare the energy densities of any solution to the Maxwell equations when looked at in the metric $g$ resp. in the metric $\bar{g}$;

$$(\text{energy density for})(\bar{g}) = \lambda(t)^4(\text{energy density for})(g).$$

Thus we have obtained another observable quantity as a function of emission time $t$:

> **Observed Luminosity for $\overline{\mathbf{g}}$**
>
> $$L_{\bar{g}} = const \cdot \mathbf{s}_\kappa(T-t)^{-2}(1+z)^{-4}.$$

The *const* depends on the telescope used. Also, my knowledge in electrodynamics is not quite sufficient to guarantee that the energy density of the solution, *computed for each frequency*, is the correct physical quantity that determines the brightness of the observed source. For example, *frequency bands* are stretched by the red shift, but this simply changes the power of $(1+z)$.
At this point one can already plot a *red shift – luminosity diagram* since both red shift and luminosity are explicit functions of the time parameter $t$. It is however instructive to eliminate $t$ from this formula and replace it by a function of $z$ to get, completely explicitly, luminosity as a function of red shift (where of course the function also depends on the two model parameters). The following formulas are used for the elimination:

Main formula to replace $t$ by $z$:

$$\lambda(t) = 1+z \quad \text{or} \quad \mathbf{s}_\kappa(t/2)^2 = \mathbf{s}_\kappa(T/2)^2(1+z)^{-1}.$$

Recall: $\mathbf{s}_\kappa'' + \kappa\mathbf{s}_\kappa = 0,\ \ \mathbf{c}_\kappa = \mathbf{s}_\kappa',\ \ \mathbf{c}_\kappa^2 + \kappa\mathbf{s}_\kappa^2 = 1.$

Functional relations:

$$\mathbf{s}_\kappa(T-t) = \mathbf{s}_\kappa(T)\mathbf{c}_\kappa(t) - \mathbf{c}_\kappa(T)\mathbf{s}_\kappa(t),$$

$$\mathbf{s}_\kappa(t) = 2\cdot\mathbf{s}_\kappa(t/2)\mathbf{c}_\kappa(t/2) = 2\cdot\mathbf{s}_\kappa(t/2)\big(1-\kappa\mathbf{s}_\kappa(t/2)^2\big)^{1/2},$$

$$\mathbf{c}_\kappa(t) = 1 - 2\kappa\mathbf{s}_\kappa(t/2)^2 = 2\mathbf{c}_\kappa(t/2)^2 - 1.$$

Model parameters:

$$\mathbf{s}_\kappa(T/2)^2 = \frac{3}{\rho(T)}, \quad \frac{1}{2q} = 1 - \frac{3\kappa}{\rho(T)} = \mathbf{c}_\kappa(T/2)^2,$$

$$H^2 = \frac{\rho(T)}{3} - \kappa = \left(\frac{\mathbf{c}_\kappa(T/2)}{\mathbf{s}_\kappa(T/2)}\right)^2 = \frac{\rho(T)}{6q}, \ \text{(see discussion above)}.$$

With these we work on the luminosity formula:

$$\mathbf{s}_\kappa(T-t) = \mathbf{s}_\kappa(T)(1-2\kappa\mathbf{s}_\kappa(t/2)^2) - \mathbf{c}_\kappa(T)2\mathbf{s}_\kappa(t/2)\mathbf{c}_\kappa(t/2)$$

$$= \mathbf{s}_\kappa(T)(1-\frac{2\kappa}{1+z}\mathbf{s}_\kappa(T/2)^2) - \mathbf{c}_\kappa(T)\frac{2\mathbf{s}_\kappa(T/2)}{\sqrt{1+z}}\left(1-\frac{\kappa\mathbf{s}_\kappa(T/2)^2}{1+z}\right)^{1/2},$$

$t$ is eliminated, next organize the model parameters to obtain a simple expresion:

$$= \frac{2\mathbf{s}_\kappa(T/2)}{1+z}\Big(\mathbf{c}_\kappa(T/2)(z+\mathbf{c}_\kappa(T)) - \mathbf{c}_\kappa(T)\big(z+\mathbf{c}_\kappa(T/2)^2\big)^{1/2}\Big)$$

$$= \frac{2\mathbf{s}_\kappa(T/2)}{1+z}\Big(\mathbf{c}_\kappa(T/2)z - (2\mathbf{c}_\kappa(T/2)^2-1)\Big(\big(z+\mathbf{c}_\kappa(T/2)^2\big)^{1/2} - \mathbf{c}_\kappa(T/2)\Big)\Big)$$

$$= \frac{2\mathbf{s}_\kappa(T/2)}{\mathbf{c}_\kappa(T/2)}\frac{z}{1+z}\left(\frac{1}{2q} - (\frac{2}{2q}-1)\Big(\big(1+2qz\big)^{1/2}+1\Big)^{-1}\right) =$$

$$= \frac{2\mathbf{s}_\kappa(T/2)}{\mathbf{c}_\kappa(T/2)} \frac{z}{1+z} \frac{\big(-1+\sqrt{1+2qz}\big)\frac{1}{2q}+1}{\sqrt{1+2qz}+1}$$

$$= \frac{2}{H} \frac{z}{1+z} \frac{\big(z+1+\sqrt{1+2qz}\big)}{(\sqrt{1+2qz}+1)^2}.$$

So, finally we have expressed $\mathbf{s}_\kappa(T-t)$ in terms of $H, q$ and $z$:

$$\mathbf{s}_\kappa(T-t) = \frac{1}{H} \frac{z}{1+z} \left(1 + \frac{(1-q)z}{1+qz+\sqrt{1+2qz}}\right)$$

For small $\kappa$ and $z$, we recognize Hubble's law, $distance \cdot H = z$. The formula is more explicit (and also dependent on the other model parameter $q$) and in this sense superior to the earlier Taylor argument. But in the present conformal description $T-t$ is not the proper time between the space slices of emission and reception, so one needs (at least for larger $T-t$) one more integration using $d\tau = dt/\lambda(t)$. – From an earlier discussion we know that "distances" are not measured directly, but are, mostly, computed from observed luminosity comparisons. We plug the expression for $\mathbf{s}_\kappa(T-t)$ into the luminosity formula and get a relation between two observable quantities (recall that the derivation is not for frequency bands, but for each frequency):

**Fully nonlinear red shift - luminosity relation**

$$L_{\overline{g}} = const \cdot \mathbf{s}_\kappa(T-t)^{-2}(1+z)^{-4}$$

$$= const \cdot \frac{H^2}{z^2}(1+z)^{-2}\left(1 + \frac{(1-q)z}{1+qz+\sqrt{1+2qz}}\right)^{-2}.$$

$const := L_{\overline{g}} \cdot distance^2$, from Hubble's law for small $z$.

So we have derived a fully nonlinear *red shift – luminosity relation* for our 2-parameter family of models. Here $H^2$ can be determined from small values of $z$ (just large enough so that the individual motions of the galaxies cause no significant errors), therefore $H^2$ is now (after the 1987 supernova) known with reasonable accuracy. For $q$ it is not so good, since for small values of $z$ the derived formula does not depend much on $q$ and for large values of $z$ the observed luminosities have large errors.

### Mass between $z$ and $z + dz$

The space slices $\{t = const\}$ of our models are orthogonal to the world lines of the dust particles and we also know $\mathrm{div}\,(\rho U) = 0$. This implies the following: if we take a ball in such a space slice, consider the world lines through all its points and follow them to another space slice, then we get again a ball and *the mass that is contained in these two balls is the same.* If we observe galaxies whose light has some precise red shift $z_*$ then this light was emitted at some time $t_*$ from a sphere in the space slice $\{t = t_*\}$. The $\overline{g}$-area $A_*$ of this sphere is $A_* = \lambda(t_*)^{-2} 4\pi^2 \mathbf{s}_\kappa(T-t_*)^2$. If we observe all galaxies with red shift between $z_*$ and $z_* + dz$ then this light was emitted from a shell of a thickness $d$ (to be computed below from $dz$) around that sphere, i.e., it comes from a volume of size $A_* \cdot d(dz)$. Therefore we see in that red shift range radiating mass of size $M_* = \rho(t_*) A_* \cdot d(dz)$. We can also *count* the number of galaxies in the red shift range $[z_*, z_* + dz]$. Either we have already determined the model parameters $H, q$ then the computed total mass and the counted number gives us the average mass of a galaxy at time $t_*$. Or else, we have some opinion about the average mass of a galaxy, then we can fit the model parameter $q$ so that the computed total mass $M_*(q)$ and the counted number of galaxies agrees with the assumed average mass. (Of course we do not need to count in the whole sky, we can select some fixed spatial angle $\Omega$ instead of $4\pi^2$.) In either case, the computation of mass per red shift range is an important model prediction. (We write $dM$ for $M_*$ and we omit all the stars.)

Since the velocity of light is 1, we have for the thickness of the spherical shell:

$$d = d\tau = \frac{dt}{\lambda(t)}, \quad \text{with} \quad 1 + z = \lambda(t) \quad \text{therefore} \quad dz = \dot{\lambda}(t) dt \quad \text{and}$$
$$d(dz) = \frac{dz}{\lambda(t)\dot{\lambda}(t)}, \quad \text{(thickness of the shell)}.$$

With $\rho(t) = \rho(T)\lambda(t)^3$ this gives:

$$dM = \frac{4\pi^2 \mathbf{s}_\kappa(T-t)^2}{\lambda(t)^2} \cdot \rho(T)\lambda(t)^3 \cdot \frac{dz}{\lambda(t)\dot{\lambda}(t)}.$$

As in the red shift – luminosity relation we have to eliminate $t$ for $z$ and organize the model parameters. $\lambda(t) = 1 + z$ is the basis of the elimination and $\mathbf{s}_\kappa(T-t)$ was already done. (Recall that for the non-realistic metric $g = g\kappa - dt^2$ the time between emission and observation is $T - t$. This is not true for the realistic metric $\overline{g} = \lambda^{-2} g$, but note that we do not need the correct time difference, we only need the conformal factor.) Therefore only $\frac{d}{dt}\lambda/\lambda$ remains:

$$\begin{aligned}
\left(\frac{\dot{\lambda}(t)}{\lambda(t)}\right)^2 &= \left(\frac{-\mathbf{s}'_\kappa(t/2)}{\mathbf{s}_\kappa(t/2)}\right)^2 = \frac{1}{\mathbf{s}_\kappa(t/2)^2} - \kappa = \frac{1+z}{\mathbf{s}_\kappa(T/2)^2} - \kappa \\
&= \frac{\rho(T)}{3}\left(z + 1 - \frac{3\kappa}{\rho(T)}\right) = \frac{\rho(T)}{3} z + H^2 \\
&= H^2(1 + 2qz), \quad \text{where} \quad \frac{\rho(T)}{3} = 2qH^2 \quad \text{was used.}
\end{aligned}$$

Now insert all the auxiliary computations into the expression for $dM$:

$$dM = \frac{4\pi^2}{H^2}\left(\frac{z}{1+z}\right)^2 \left(1 + \frac{(1-q)z}{1 + qz + \sqrt{1+2qz}}\right)^2 \cdot \frac{\rho(T)}{1+z} \cdot \frac{dz}{H\sqrt{1+2qz}}$$

After one more simplification - use $\rho(T) = 6qH^2$ - we obtain the promised

**Mass per red shift relation**

$$dM = \frac{24\,\pi^2}{H} \frac{z^2 dz}{(1+z)^3} \frac{q}{\sqrt{1+2qz}} \left(1 + \frac{(1-q)z}{1+qz+\sqrt{1+2qz}}\right)^2 .$$

Recall that $H$ is known with good accuracy and observe that this relation varies much more with $q$ than the previous one so that one does not have to rely on large values of $z$ to see the $q$-dependence in the model prediction. I expect that, for smaller $z$, astronomers have reasonable estimates for the average mass of a galaxy so that our result can be used to determine $q$ (as the second model parameter) from galaxy counts in the red shift range $[z, z+dz]$.

## Problems with Gravitational Waves

The basic questions are: what do we expect a gravitational wave to be? and what observational effects do we expect such a wave to have?

In the well understood electromagnetic case we have the electric and magnetic fields, represented by the Faraday 2-form $F$, we have a pair of first order differential equations for $F$ (Maxwell's) and any such field (wave or not) leads to a covariant acceleration of the world lines $\gamma$ of charged particles, given by the Lorentz force: $\frac{D}{ds}\gamma' = (e/m)\widehat{F}(\gamma')$.

In the gravitational case we need intuition building analogies. Consider first the solar system. We compared the eigenvalues of the Hessian of the Newtonian potential with the eigenvalues of the curvature along a planetary world line $\gamma$, i.e. the eigenvalues of $R(\,.\,,\gamma')\gamma'$. Up to relativistic correction factors like $(1-2m/r)$ they turned out to be the same: $-2m/r^3, m/r^3, m/r^3$. Therefore one has to view the Hessian and the curvature as corresponding objects. This leads one to view the Newtonian gravitational field and the Schwarzschild Christoffel symbols as analogous objects and, one more anti-derivative up, the Newtonian potential and the Schwarzschild metric are viewed as analogous.
The same conclusion is reached in the following weak field situation:
Consider a metric $g = (1-2\Phi(x,y,z))(dx^2+dy^2+dz^2) - (1+2\Phi(x,y,z))dt^2$. Here the spatial part of the geodesic equation takes the form: $(x''(s), y''(s), z''(s)) \approx -\mathrm{grad}\,\Phi \cdot t'(s)^2$. Now the assumption "weak field" means that $t'(s)^2 \approx 1$. One can summarize this by saying: The relativistic equation of motion, i.e. the geodesic equation, equals – in the special coordinates in which the metric is written – the Newtonian equation of motion, so that again the Christoffel symbols are found to be the relativistic rendering of the Newtonian forces, of $\mathrm{grad}\,\Phi$. (These weak field computations are a bit strange: The assumptions are so restrictive that no relativistic effects are dealt with. I do not remember seeing the stress energy tensor discussed. Its leading spatial part is the Hessian of $\Phi$, which looks more like non-isotropic elastic material than having to do with gravity. The sole purpose seems to be to embed a Newtonian situation into a relativistic setting so that in one special coordinate system one has $\gamma'' = -\mathrm{grad}\,\Phi$ and proper time pretty equal to coordinate time.)

These analogies led to the following widely accepted definition of the notion of "gravitational wave": Consider the Einstein equations as a second order PDE for the metric. Linearize the Einstein equations along the flat Minkowski metric of Special Relativity and consider this linearized equation as the gravitational wave equation. Notice that one obtains uninteresting solutions as follows: Pull back the metric by a group of diffeomorphisms and linearize this family. On the one hand one obtains a solution of the linearized Einstein equations, on the other hand, a pull back of the metric only changes the description of the geometry, not the geometry itself. Such solutions are called coordinate waves and are eliminated by a gauge procedure. Next we derive (for an arbitrary background metric $g$) the **Linearized Einstein Equations:**

$h(Y,Z) := \frac{d}{d\epsilon} g_\epsilon(Y,Z))$ Linearization of the metric, $g_0 = g$

$D^\epsilon_Y Z = D_Y Z + \Gamma_\epsilon(Y,Z)$ Difference tensor of covariant derivatives

$\gamma(Y,Z) := \frac{d}{d\epsilon}\Gamma_\epsilon(Y,Z)$ Linearization of the Christoffel symbols

Relations between $Dh$ and $\gamma$:

$$(D_X h)(Y,Z) = g(\gamma(X,Y),Z) + g(Y,\gamma(X,Z))$$

$$g(\gamma(Y,Z),X) = \frac{1}{2}\big( -(D_X h)(Y,Z) + (D_Y h)(Z,X) + (D_Z h)(X,Y)\big)$$

$$g((D_X\gamma)(Y,Z),W) = \frac{1}{2}\big( -(D^2_{X,W} h)(Y,Z) + (D^2_{X,Y} h)(Z,W) + (D^2_{X,Z} h)(W,Y)\big)$$

Linearized curvature tensor $linR$ and linearized Ricci tensor $linric$:

$$linR(X,Y)Z := \frac{d}{d\epsilon} R_\epsilon(X,Y)Z = (D_X\gamma)(Y,Z) - (D_Y\gamma)(X,Z)$$

$$\begin{aligned} linric(Y,Z) &= \text{trace}\,(X \mapsto linR(X,Y)Z) \\ &= \text{trace}\,(X \mapsto (D_X\gamma)(Y,Z) - (D_Y\gamma)(X,Z)) \\ &= \sum_i \big(g((D_{e_i}\gamma)(Y,Z),e_i) - g((D_Y\gamma)(e_i,Z),e_i)\big)/g(e_i,e_i) \end{aligned}$$

We are after the equation away from sources, analogous to the homogenous Maxwell equation $d{*}F = 0$. This means that the linearized stress energy tensor is zero and therefore the linearized Einstein tensor, $linG$, is equal to the linearized Ricci tensor. To keep the formula more readable we do not insert for $D\gamma$ its expression in terms of $D^2h$, although we want the result to be understood as a second order equation for $h$:

**Linearized Einstein equation**, insert $D^2h$ for $D\gamma$

$$0 = \text{trace}\,(X \mapsto (D_X\gamma)(Y,Z) - (D_Y\gamma)(X,Z))$$

In the literature this equation is, in the case of a **flat** background metric $g$, discussed as the equation describing **gravitational waves**. A gravitational wave then "is" the symmetric

2-tensor field $h$.

As an **example** of such a linearization we consider the family of Schwarzschild metrics

$$g_m = (1 - 2m/r)^{-1} dr^2 + r^2 d\sigma^2 - (1 - 2m/r) dt^2$$

Clearly $g_0$ is the flat metric of Special Relativity. We compute the linearization:

$$g_0 + m \cdot \frac{d}{dm} g_m|_{m=0} = (1 + 2m/r) dr^2 + r^2 d\sigma^2 - (1 - 2m/r) dt^2.$$

Noting that $-m/r$ is the Newtonian potential we take from this result the suggestion for the weak field Ansatz above. The Schwarzschild linearization is a sufficiently good approximation to compute the deflection of light near the sun, but the linearization does not give the correct value of the perihelion advance since the "relativistic correction factors" $(1 - const \cdot m/r)$ do not come out right. If one could find a similar solution on the Schwarzschild geometry itsself such that the singularity has a world line like a planet, then one would be a step closer to a relativistic description of a 2-body problem.

Maybe it is useful to have seen **coordinate waves**. Let $X$ be a vector field and $\Psi_\epsilon$ be its flow, so that $\frac{d}{d\epsilon}\Psi_\epsilon(p) = X(p)$ and $\frac{D}{d\epsilon}(T_U \Psi_\epsilon) = D_U X$. Linearize the family of pull back metrics $g_\epsilon(V, W) := g(T\Psi_\epsilon V, T\Psi_\epsilon W)$ so that

$$\begin{aligned} h(V, W) &:= \frac{d}{d\epsilon} g_\epsilon(V, W) = g(D_V X, W) + g(V, D_W X) \\ (D_U h)(V, W) &= g(D^2_{U,V} X, W) + g(V, D^2_{U,W} X) \\ &= g(\gamma(U, V), W) + g(V, D\gamma(U, W)). \end{aligned}$$

By the symmetry $\quad (D^2_{U,V} X + R(X, U)V) \; - \; (D^2_{V,U} X + R(X, V)U) = 0$

and $\quad g(R(X, U)V, W) + g(V, R(X, U)W) = 0$

we get $\quad \gamma(U, V) = D^2_{U,V} X + R(X, U)V.$

This implies for the linearized curvature of $h$:

$$\begin{aligned} linR(U, V)W &= (D_U \gamma)(V, W) - (D_V \gamma)(U, W) \\ &= -D_{R(U,V)W} X + (D_X R)(U, V)W \\ &\quad + R(D_U X, V)W + R(U, D_V X)W + R(U, V) D_W X. \end{aligned}$$

In the textbook situation of a **flat** background metric $g$ the coordinate waves cannot have curvature either. In the general case I do not think that one can easily recognize whether $h$ is a coordinate wave since along two different timelike geodesics the curvature of $g$ will in general be different and therefore $linR \neq 0$.

Next we discuss what **observational effects** a gravitational wave might have. One may compute predictions in two different geometries, but there is no way in which an experiment could be made that *measures* a difference of what happens in one geometry (one universe) and in a second geometry. As long as one has only one observer on a single world line, this

world line is a geodesic; as long as one does not consider another (nearby or not) world line there is no observable difference between different geometries.
As soon as we observe **two** nearby world lines we can observe the relative acceleration, the second covariant derivative of the separation vector. It satisfies the Jacobi equation $\frac{D}{ds}(\frac{D}{ds}J) + R(J,\gamma')\gamma' = 0$. All gravitational wave detectors whose design descriptions I have understood are capable of measuring such relative accelerations. What can one hope to measure? The curvature of the Friedmann universes discussed above is too small to be measured via the acceleration of separation vectors any time soon. But the Friedmann universes are highly simplified models, in the real world supernova explosions do occur, heavy stars, maybe even heavier black holes, rotate around each other. This will be reflected in the curvature tensor, also along our own world line. The expectation therefore is that the relevant curvature in the above Jacobi equation is not the exceedingly small Friedmann curvature, but are much larger time dependent curvature fluctuations. As soon as the instruments become sensitive enough, such fluctuations will become measurable. From that moment on will the discussion, whether such fluctuations are described by the above or any other wave equation, be *influenced by observations* - certainly Maxwell's equations were only formulated after very careful and surprising experiments.
Summary:

Electromagnetic waves (covariantly) accelerate charged particles.
Gravitational waves (covariantly) accelerate separation vectors of particle pairs.
On rest spaces $\gamma'^{\perp}$ of observers $\gamma$ a gravitational wave acts as a symmetric 2-tensor $x \mapsto R(x,\gamma')\gamma'$. Its trace is $ric(\gamma',\gamma')$ (the trace of the linearized curvature is 0).

Finally I discuss the problems that I have with the linearized Einstein equations.

1.) Gregor Weingart proved that those equations, when considered on a **non-flat** background, do not have enough solutions. But for a wave equation one should be able to pose initial value problems. For some PDEs, for example the equation for Killing vector fields, $DX + DX^{transpose} = 0$, it is well known that a complicated curvature tensor restricts or even prevents solutions. The linearized Einstein equation is not among the well known cases and I cannot sketch Gregor's proof either. But the following suggests that one should want to see a proof before one believes that the linearized Einstein equation on a non-flat background behaves just like a wave equation. It is convenient, to represent $h$ by a symmetric endomorphism field $H$ as $h(Y,Z) = g(H\cdot Y,Z)$. First we have the usual relation of the second derivative of a tensor field with the curvature:

$$(D^2_{U,V}h - D^2_{V,U}h)(Y,Z) = -h(R(U,V)Y,Z) - h(Y,R(U,V)Z) = g(\big[R(U,V),H\big]Y,Z).$$

On the other hand we have the relation with the linearized curvature:

$$\begin{aligned}&(D^2_{U,V}h - D^2_{V,U}h)(Y,Z)\\ &= g((D_U\gamma)(V,Y),Z) + g(Y,(D_U\gamma)(V,Z)) - g((D_V\gamma)(U,Y),Z) - g(Y,(D_V\gamma)(U,Z))\\ &= g(linR(U,V)Y,Z) + g(Y,linR(U,V)Z).\end{aligned}$$

Hence $\big[R(U,V),H\big] = linR(U,V) + linR(U,V)^{transpose}$.

This result says that for sufficiently general curvature $R(U,V)$ the trace free part of $H$ is *algebraically* determined by the linearized curvature $linR$ of $H$. Now the linearized curvature is restricted by the linearized Einstein equation, $\mathrm{trace}\,(x \mapsto linR(x,V)W) = 0$. Therefore also the values of $H$ are restricted. This does **not** happen for a **flat** background metric $g$ and is therefore not part of the gravitational wave discussion, but it is clearly an unwanted feature of a "wave" equation.

2.) The above discussion about observations explained that fluctuations such as the intuitively expected waves should make themselves felt as relative acceleration between two world lines. From a differential geometric point of view it is very strange that an agreement was reached saying that the forces caused by gravitational waves should be described by the Christoffel symbols. These Christoffel symbols by themselves do not have an invariant meaning, and for the same reason, the second coordinate derivatives of the world lines have no invariant meaning, in particular they are not accelerations of world lines. Recall that the forces of electromagnetism, the forces of this experimentally tested theory, cause a **covariant** acceleration of the world lines of the charges.

3.) It is easy to misinterpret this discussion. I am definitely **not** saying: "How can they try to measure things that do not exist?" For example the computation of perihelion advance in the Schwarzschild geometry predicts an acceleration of separation vectors (caused by curvature) that turned out to be measurable. Similarly, with sufficiently sensitive apparatus we will be able to measure relative accelerations that are not caused by anything in the solar system, simply because the curvature tensor along the world lines of observers is not determined by the solar system. However, this does not imply that we can separate, say, supernova caused fluctuations in such a way from a simpler background geometry $g$ that we can *describe* these fluctuations as solutions of some linear PDE, either as solutions of the linearized Einstein equations or of other suggested equations, on the background geometry $g$.