IOP Concise Physics | A Morgan & Claypool Publication

Special and General Relativity

An Introduction to Spacetime and Gravitation

Rainer Dick

IOP | ebooks

Special and General Relativity

An introduction to spacetime and gravitation

Special and General Relativity

An introduction to spacetime and gravitation

Rainer Dick
Department of Physics and Engineering, University of Saskatchewan, Saskatoon, Canada

Morgan & Claypool Publishers

ISBN 978-1-64327-380-8 (ebook)
ISBN 978-1-64327-377-8 (print)
ISBN 978-1-64327-378-5 (mobi)

DOI 10.1088/2053-2571/aaf173

Version: 20190201

IOP Concise Physics
ISSN 2053-2571 (online)
ISSN 2054-7307 (print)

A Morgan & Claypool publication as part of IOP Concise Physics
Published by Morgan & Claypool Publishers, 1210 Fifth Avenue, Suite 250, San Rafael, CA, 94901, USA

IOP Publishing, Temple Circus, Temple Way, Bristol BS1 6HG, UK

For Josephin, Fabian and Jonathan

Contents

Preface

The inception of relativity as a dynamical theory of spacetime was a major breakthrough of curiosity driven basic research, and it had important ramifications both for particle physics and for our understanding of the structure and evolution of the Universe on the largest scales. Training in special relativity is therefore an integral part of the education of every aspiring physicist, and training in general relativity is indispensable for every student who wishes to pursue graduate studies in theoretical physics, astrophysics, or cosmology.

Furthermore, only about half a century after the inception of relativity, the technological relevance of relativity emerged in the analysis of satellite signals and orbits. This was not noticed by the public yet, and both special and general relativity continued to be widely perceived as interesting theories with significant scientific impact, but little *practical* relevance. This perception started to change about three decades later, in the mid-1980s, when the global positioning system (GPS) became available for civil purposes and more people learned about the immediate relevance of relativity for the operation of GPS.

The wide ranging scientific implications and the modern technological relevance of relativity imply that learning relativity is one of the highlights in the training of physics students. In the North American physics curriculum this usually proceeds in two or three stages: a first introduction to relativistic mechanics is often provided as part of a second year course on the foundations of modern physics, while aspects of relativistic electrodynamics are discussed as part of a third or fourth year course on electromagnetism. This usually concludes the mandatory training in relativity for physics students and is limited to special relativity. A deeper introduction to special relativity along with an introduction to general relativity is then provided through a fourth year elective course or an introductory graduate course. The present book should be helpful at every stage of this traditional three-step approach to general relativity. However, it should also help to accelerate and streamline the process of learning relativity by unifying the introduction to special and general relativity in a single concise book, which can be used as a basis of a one-semester course taking students all the way from the foundations of special relativity to the derivation of the Schwarzschild metric and particle motion in general relativity. Furthermore, the immediate impact of relativity for time-keeping and signal transmission would make it prudent for engineering colleges to include introductory relativity courses as electives for students in electrical, computer, and aerospace engineering. It is therefore timely to make the theory of relativity more easily accessible to all students in science and engineering. This book intends to serve that purpose through its concise presentation and by making special relativity accessible to first year students, while the general relativity part should be understandable to third year students.

The book evolved from notes for the special relativity sections of my second year modern physics course, and from the lecture notes for my course on general relativity, gravitation, and cosmology. The latter course is offered biannually for third or fourth year undergraduate students and beginning graduate students.

The first three chapters do not require any preparation beyond a standard first year physics course. They can therefore be used as a reference for a first introduction to special relativity towards the end of the first year physics course, or as part of the second year modern physics course. Students who learn special relativity as part of their third or fourth year electromagnetism courses should also find these chapters useful.

Chapters 4–9 require that students have taken second year courses on vector calculus and mechanics, and second or third year courses on partial differential equations and electrodynamics. In North American universities, students in physics and engineering have usually acquired the necessary skills in vector calculus, intermediate mechanics, and partial differential equations after the first term of their third year, and they usually also take an electrodynamics course during their third year. The present book can therefore serve as a textbook for a course which is already offered in the second term of the third year, thus providing both students and instructors with a lot of flexibility whether they want to learn or teach general relativity in third year, fourth year, or at the introductory graduate level.

Author biography

Rainer Dick

Rainer Dick studied physics at the Universities in Stuttgart, Karlsruhe and Hamburg, and received a PhD degree from the University of Hamburg in 1990. He worked at the University of Munich and the Institute for Advanced Study in Princeton before joining the University of Saskatchewan in 2000. Rainer's research interests span a wide range of topics from particle physics, cosmology and string theory to materials physics and quantum optics. Rainer has published over 100 papers in journals and conference proceedings, and a textbook on *Advanced Quantum Mechanics: Material and Photons*.

Chapter 1

Why relativity?

1.1 The Galilei invariance of Newtonian mechanics

Before we learn relativity, we all live our lives with an implicit *prejudice:* In agreement with our everyday experience, *we assume that every person always measures the same time* (up to conversions between time zones, but let's just assume that we all use Greenwich time). This prejudice can also be denoted as an *assumption of universal time*. If an astronaut on the space station and an astronomer on Earth observe the brightening and subsequent fading of a supernova, then we naively think that they would assign the same time interval between the occurrence and the fading of the supernova, and this point of view was in agreement with the science of mechanics for over two centuries.

To explain the connection with mechanics, we need to recognize that the *assumption of universal time* implies certain coordinate transformation equations between observers which move at constant velocity $\boldsymbol{v}$ relative to each other: Suppose two observers, Frank and Mary, each use their own coordinate system $\{x, y, z\}$ (Frank) and $\{x', y', z'\}$ (Mary), to describe the 'event' in figure 1.1. We also assume that Frank and Mary met at time $t = t' = 0$ and that Mary moves with velocity $\boldsymbol{v} = (v, 0, 0) = v\boldsymbol{e}_x$ relative to Frank, where $\boldsymbol{e}_x$ denotes the unit vector in the x-direction. The assumption $t' = t$ then implies that we should be able to read off from figure 1.1 that Frank's and Mary's coordinates for the event are related according to

$$x' = x - vt,\ y' = y,\ z' = z, \tag{1.1}$$

because after time $t = t'$ elapsed, Mary has moved a distance $vt = vt'$ away from Frank in the x-direction, such that her x'-coordinate to the event is smaller than Frank's x-coordinate by the amount vt.

This simple example illustrates already that the world is four-dimensional, even without relativity: We need four coordinates t, x, y, z to localize an event in space and time. If we explicitly remind ourselves that we use the assumption of universal

doi:10.1088/2053-2571/aaf173ch1

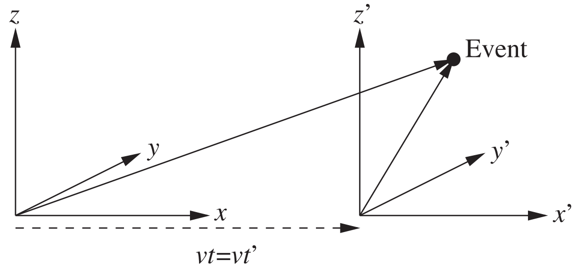

Figure 1.1. Two frames with relative velocity $\boldsymbol{v} = (v, 0, 0) = v\boldsymbol{e}_x$ and the assumption of universal time: $t = t'$.

time, the transformation equations (1.1) become the *Galilei transformation* for relative motion with velocity $\boldsymbol{v} = v\boldsymbol{e}_x$,

$$t' = t,\ x' = x - vt,\ y' = y,\ z' = z. \tag{1.2}$$

However, it is crucial to recognize that our inference of the equations (1.1) from figure 1.1 used the assumption of universal time, $t' = t$, in two ways: The assumption $t = t'$ implies the further assumption that both observers see the *same momentary picture* 1.1 when the times t and t' have elapsed on their clocks, and therefore we also infer a third assumption, viz that both observers assign the *same distance vector* with length $vt = vt'$ to their separation at times t and t'. However, all this is wrong! Once we are forced to surrender the assumption of universal time, we also have no more reason to believe that Frank's perception of the two frames at time t equals Mary's perception at time t', and the distance vectors $\boldsymbol{v}t$ and $\boldsymbol{v}t'$ will also not be the same any more. Therefore a basic mistake in figure 1.1 is the inclusion of the equation $vt = vt'$ and the corresponding identification of the distance vectors at times t and t'. We will discuss this in detail in chapter 3.

However, for now let's stick with our naive assumption $t = t'$ and see how this seemed to be confirmed by Newtonian mechanics.

Suppose the 'event' in figure 1.1 is a moving particle of mass m at time t. If this particle moves under the influence of a force $\boldsymbol{F}(\boldsymbol{x}, t)$ (e.g. due to a local electric field which also changes with time), then Newtonian mechanics says that the particle satisfies the equation of motion

$$m\frac{d^2}{dt^2}\boldsymbol{x}(t) = \boldsymbol{F}(\boldsymbol{x}(t), t) \tag{1.3}$$

in the $\{t, \boldsymbol{x}\}$ frame. However, the Galilei transformation

$$t' = t,\ \boldsymbol{x}'(t') = \boldsymbol{x}(t) - \boldsymbol{v}t, \tag{1.4}$$

into the $\{t', \boldsymbol{x}'\}$ frame then implies that the particle also satisfies Newton's equation in that frame with a local force $\boldsymbol{F}'(\boldsymbol{x}'(t'), t')$,

$$
\begin{aligned}
m\frac{d^2}{dt'^2}\boldsymbol{x}'(t') = m\frac{d^2}{dt^2}\boldsymbol{x}(t) = \boldsymbol{F}(\boldsymbol{x}(t), t) = \boldsymbol{F}(\boldsymbol{x}'(t') + \boldsymbol{v}t', t') \\
= \boldsymbol{F}'(\boldsymbol{x}'(t'), t'),
\end{aligned}
\tag{1.5}
$$

with the transformation law for forces between the two reference frames,

$$
\boldsymbol{F}'(\boldsymbol{x}', t') = \boldsymbol{F}(\boldsymbol{x}, t). \tag{1.6}
$$

If we also include a rotation $\underline{R}$ of the coordinate axes and constant coordinate shifts T and $\boldsymbol{X}$ (which implies $\boldsymbol{v}' = \underline{R} \cdot \boldsymbol{v}$ and $\boldsymbol{X}' = \underline{R} \cdot \boldsymbol{X}$ for the relative velocity components and spatial shifts assigned by the two observers),

$$
t' = t - T,\ \boldsymbol{x}'(t') = \underline{R} \cdot (\boldsymbol{x}(t) - \boldsymbol{v}t - \boldsymbol{X}),\ \boldsymbol{x}(t) = \underline{R}^T \cdot (\boldsymbol{x}'(t') + \boldsymbol{v}'t' + \boldsymbol{X}'), \tag{1.7}
$$

then equations (1.5) and (1.6) become

$$
m\frac{d^2}{dt'^2}\boldsymbol{x}'(t') = \boldsymbol{F}'(\boldsymbol{x}'(t'), t') \tag{1.8}
$$

and[1]

$$
\boldsymbol{F}'(\boldsymbol{x}', t') = \underline{R} \cdot \boldsymbol{F}(\boldsymbol{x}, t). \tag{1.9}
$$

Equation (1.8) with the constant coordinate shifts T and $\boldsymbol{X}$ included is denoted as an *inhomogeneous Galilei transformation*, whereas the transformation (1.4) is denoted as a *Galilei boost*.

Equations (1.3) and (1.8) express the *form invariance* (or *invariance* for short) of Newton's equation under Galilei transformations. If Newtonian mechanics holds in a reference frame $\{t, \boldsymbol{x}\}$, then it also holds in every reference frame $\{t', \boldsymbol{x}'\}$ which relates to the frame $\{t, \boldsymbol{x}\}$ through a Galilei transformation (1.7).

How did physicists then learn that, in spite of the invariance of Newtonian mechanics under Galilei transformations, the assumption of universal time is wrong? We will not explain relativity theory in the remainder of this chapter (that requires a bit more preparation and has to wait until chapter 3 and the following chapters), but we will review the observations and developments which made relativity unavoidable in sections 1.2 and 1.3.

1.2 The need for special relativity

There were theoretical and experimental observations in the 1880s that something was amiss with Galilei transformations:

1. The laws of electromagnetism are not invariant under Galilei transformations, i.e. if Maxwell's equations hold in the $\{t, \boldsymbol{x}\}$ coordinate system, they

[1] We will learn in due course that a transformation law $F'(x') = F(x)$ of a quantity $F(x)$ under coordinate transformations $x \to x'$ is denoted as a *scalar* transformation. Equation (1.6) tells us that forces transform like scalars under Galilei boosts (1.4). However, equation (1.9) tells us that forces transform like *vectors* under the general Galilei transformation (1.7).

would not hold in the $\{t', \boldsymbol{x}'\}$ coordinate system if the coordinates are related by a Galilei transformation[2] (1.7).

This is in striking difference to the Galilei invariance of Newton's equation: For Newtonian mechanics the two frames are completely equivalent, but for electrodynamics they are inequivalent!

2. If Galilei transformations would be a correct description of the coordinate transformations, Frank and Mary would observe different speeds for light fronts:

   Suppose Frank triggers a light flash. In his frame this will generate a spherical light front moving with speed c in all directions.

   For Mary this light front would move with speed $c - v$ in x-direction, with speed $c + v$ in $(-x)$-direction, and with speed c orthogonal to the x-axis, if the transformation equations (1.2) between their coordinates were correct. *However, this prediction was shown to be false in the Michelson experiment*[3].

Both of these problems were solved by Einstein's inception of the special theory of relativity[4] (STR).

1. STR explained the unexpected and very puzzling outcome of the Michelson experiment. Michelson had observed that the Earth's motion around the Sun at a speed of about 30 km s^{-1} does not affect the speed of light observed in a terrestrial laboratory, irrespective of the direction in which the light wave travels. Although Einstein seemed to have been primarily motivated by the symmetry properties of electrodynamics, his work solved the Michelson problem by implying that different observers measure time and space intervals in such a way that they both find the same value c for the speed of a light wave in vacuum, irrespective of their mutual relative velocity $\boldsymbol{v}$. Specifically, this implies that space and time intervals in one coordinate system satisfy $c^2\Delta t^2 = \Delta \boldsymbol{x}^2$ if and only if they satisfy in another coordinate system the condition $c^2\Delta t'^2 = \Delta \boldsymbol{x}'^2$, if the other coordinate system moves with a constant velocity $\boldsymbol{v}$ relative to the first system. For relative velocity in x-direction, $\boldsymbol{v} = (v, 0, 0) \equiv v\boldsymbol{e}_x$, this yields transformation laws for the coordinate intervals and coordinates

$$\Delta x' = \gamma(\Delta x - \beta c\Delta t), \quad \Delta y' = \Delta y, \quad \Delta z' = \Delta z, \tag{1.10}$$

$$c\Delta t' = \gamma(c\Delta t - \beta\Delta x), \tag{1.11}$$

and

---

[2] Voigt W 1887 *Nachrichten von der Königlichen Gesellschaft der Wissenschaften und der Georg-Augusts-Universität zu Göttingen* **1887** (2) 41. This paper analyzes the symmetries of the electromagnetic wave equations and essentially contains the first incarnation of the Lorentz transformations (up to a scaling transformation).

[3] Michelson A A and Morley E W 1887 *Am. J. Sci.* 34 427; Michelson A A and Morley E W 1887 *Philos. Mag.* 24 463.

[4] Einstein A 1905 *Ann. Phys.* 17 891.

$$x' = \gamma[x - X - \beta c(t - T)], \quad y' = y - Y, \quad z' = z - Z, \tag{1.12}$$

$$ct' = \gamma[c(t - T) - \beta(x - X)], \tag{1.13}$$

where $\beta = v/c, \gamma = (1 - \beta^2)^{-1/2}$ and X, Y, Z, T are constant coordinate shifts.

Einstein's work provided an important re-interpretation and logical completion of previous work by Lorentz and Poincaré on the Michelson experiment, and the transformations (1.10), (1.11) and (1.12), (1.13) are known as Lorentz and Poincaré transformations, respectively.

Einstein's reasoning implied that the vacuum speed of light c is absolute, but time is not absolute! The relativity of time for different observers, and the corresponding relativity of the notions of space and time, provided the motivation for the name *relativity theory* for the new theory.

2. STR explained why the wave equations derived from Maxwell's equations are invariant under Lorentz transformations instead of the classically expected Galilei transformations. It was then demonstrated by Minkowski a few years after the invention of STR that the Lorentz transformations are in fact symmetries of the full Maxwell's equations[5].

1.3 The need for general relativity

Explaining the universality of the vacuum speed of light and the symmetries of Maxwell's equations were outstanding successes, and yet Einstein had to face two new challenges as a consequence of STR:

1. While STR explained the symmetry properties of Maxwell's equations, the very successful theory of Newtonian gravity was now incompatible with the basic symmetry principles of the theory.
2. The problem of 'proper' and 'improper' coordinate frames, which had been around in mechanics ever since the publication of Newton's *Principia*, now appeared in a different guise: STR had to assume that there exists a fixed four-dimensional flat spacetime and 'inertial frames', i.e. coordinate systems which move uniformly through this spacetime. The basic equations of STR and electrodynamics were assumed to hold in these distinguished coordinate systems, and the challenge was to identify these special coordinate systems.

For the latter problem physicists ever since Newton referred to the rest frame of the 'fixed stars' as an example of an inertial frame, and any other frame that was related to this frame through Galilei transformations (mechanics) or Lorentz transformations (STR) was also considered to be an inertial frame.

In modern terms the rest frame of the 'fixed stars' would rather be denoted as the *CMB frame* (cosmic microwave background frame) or the *cosmic rest frame*.

[5] Minkowski H 1908 *Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse* **1908** (1) 53.

Nevertheless, before the inception of the general theory of relativity (GTR) we encounter two (supposedly equivalent, but actually inequivalent) definitions of inertial frames:

1. The cosmic rest frame and any non-rotating frame which moves with constant velocity $\boldsymbol{v}$ relative to the cosmic rest frame are inertial frames.
2. A coordinate system in which every trajectory $\boldsymbol{x}(t)$ of a free particle satisfies

$$\frac{d^2\boldsymbol{x}}{dt^2} = 0, \tag{1.14}$$

is an inertial frame. In different terms: the motion of a free particle is inert in an inertial frame (hence the designation 'inertial frame').

The supposed (or naively presumed) equivalence of these definitions assumes that free particles satisfy equation (1.14) in the cosmic rest frame. However, the Universe expands, and therefore the cosmic rest frame turns out *not* to be an inertial frame, because free particles do *not* satisfy an equation like (1.14) in this frame[6].

The second definition by itself, without reference to a specific example of an inertial frame, is actually meaningless unless we can find a definition of 'free particle' which does not require the notion of inertial frames to start with (if we want to use (2) as a definition of inertial frames, we better not try to define free particles with a statement like 'free particles are particles which have constant velocity, i.e. $d^2\boldsymbol{x}/dt^2 = 0$, in an inertial frame').

General relativity resolved this whole conundrum with the definition and presumed importance of inertial frames in the following way:

1. Inertial frames are actually not required for the proper formulation of the laws of nature. With a proper understanding of differential geometry, the laws of nature assume the same form in any coordinate system.
2. As a consequence of **1**, we can give a definition of free particles which works in every coordinate system and does not rely on the notion of inertial frames.
3. Indeed, inertial frames can be defined as coordinate systems where the equation of motion of free particles takes the simple form (1.14), but these inertial frames turn out to have limited extension in spacelike directions.

In a nutshell, general relativity provided a way to define free particles without reference to inertial frames, but at the same time also deprived inertial frames of their prominent role and identified limitations to the construction of inertial frames.

[6] In more precise terms, there is no global coordinate system for the Universe where all free particles satisfy an equation like (1.14) everywhere at all times. However, if we study only phenomena on time scales which are very small compared to the lifetime $t_0 \simeq 13.8$ billion years $\simeq 4.35 \times 10^{17}$ s of the Universe (i.e. if we do anything but large scale astronomy or cosmology), then a corresponding spacelike slice through the cosmic rest frame with small duration $\Delta t \ll t_0$ provides a very good approximation to an inertial frame as long as we stay away from neutron stars or black holes.

The equivalence of inertial and gravitational mass helped Einstein to solve the problem to reconcile gravity and relativity[7]. To understand this equivalence, recall that Newton[8] had solved the problem to find a dynamical explanation of Kepler's laws of planetary motion (Kepler 1609, 1619) by relating forces to acceleration, and by assuming his famous force law for the gravitational forces between masses m and M at locations $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$:

$$m\frac{d^2}{dt^2}\boldsymbol{r}_1 = -M\frac{d^2}{dt^2}\boldsymbol{r}_2 = G\frac{mM}{|\boldsymbol{r}_1 - \boldsymbol{r}_2|^2}\frac{\boldsymbol{r}_2 - \boldsymbol{r}_1}{|\boldsymbol{r}_2 - \boldsymbol{r}_1|}. \tag{1.15}$$

This contains a peculiarity which appears as a mere coincidence in classical mechanics: The acceleration $d^2\boldsymbol{r}_1/dt^2$ of the mass m is independent of m, since the 'inertial mass' m in $\boldsymbol{F} = m\boldsymbol{a}$ cancels the 'gravitational mass' m in $GmM/|\boldsymbol{r}_1 - \boldsymbol{r}_2|^2$.

Einstein realized in a famous 'Gedankenexperiment' ('thought experiment') that this implies a striking similarity between gravitational forces and inertial forces (pseudo-forces). Assume an observer standing in an elevator at rest in a homogeneous gravitational field, with the gravitational acceleration $\boldsymbol{g}$ pointing downwards. If the observer releases two different objects with different masses, both objects will hit the floor of the elevator at the same time, since they experience the same acceleration $\boldsymbol{g}$. Now assume that the elevator is far away from gravitational fields, but accelerated upwards with acceleration $\boldsymbol{a} = -\boldsymbol{g}$. Again the observer releases the two objects, which now *appear* to be accelerated downwards with acceleration $\boldsymbol{g}$, and again would hit the elevator floor simultaneously.

The outcome in both experiments is identical, and the observer cannot decide whether the force acting on the two objects was due to an external homogeneous gravitational field, or whether the force was an inertial force due to the observer's accelerated motion.

Einstein concluded that *locally we cannot separate gravitational forces from inertial forces*. But if gravitational forces are so similar to pseudo-forces, *maybe they should not be considered as genuine forces at all! Maybe what we perceive as motion of particles under the influence of gravitational forces is actually only the motion of free particles through the curved geometry of spacetime!*

The stage was set for the invention of general relativity: Gravity is a manifestation of the geometry of spacetime.

Note that the equivalence of inertial and gravitational mass plays a key role in Einstein's Gedankenexperiment: If inertial and gravitational mass were different, then two different objects would experience *different* accelerations in a homogeneous gravitational field, but they would still appear to undergo the *same* acceleration $-\boldsymbol{a}$ relative to an accelerated coordinate frame. The observer could then unambiguously distinguish between gravitational forces and inertial forces.

[7] For a historical account of the development of mechanics, see Mach E 1902 *The Science of Mechanics: a Critical and Historical Account of Its Development* 2nd edn (Chicago: Open Court Publishing Company) (Engl. transl. T J McCormack).

[8] Newton assembled all the pieces of the puzzle in several works from 1666 to 1687. This is also discussed in Mach E 1902 loc. cit.

Einstein's identification of motion in 'gravitational fields' as free motion through curved spacetime explains the coincidence of inertial and gravitational mass: The mass of a particle should have no impact on its force free motion, and therefore it should have no effect on the free motion which we perceive as motion in a gravitational field. And we will also see that it naturally solved the problems which Einstein was facing as a consequence of STR: The need to extend STR to a geometric theory of a curved spacetime automatically incorporates gravity, and it deprives inertial frames of their prominent role.

We will see this explicitly in chapters 7–9. However, first we need to acquire a better understanding of geometry and of special relativity.

Chapter 2

A first look at notions from geometry

We need a better understanding of geometry both for special relativity and for general relativity. For the time being we will start with some simple but useful definitions, and we will use the simple example of the two-dimensional Euclidean plane as a playground.

2.1 Vectors and tensors

If we consider a particle moving in a plane, we usually assign several two-dimensional vectors to it, like the vector $\boldsymbol{r}$ of the location of the particle relative to some arbitrarily chosen fixed point in the plane, and the corresponding velocity and acceleration vectors $\boldsymbol{v} = d\boldsymbol{r}/dt$ and $\boldsymbol{a} = d^2\boldsymbol{r}/dt^2$. All these vectors come with their specific dimensions like meter, meter/second, etc, but the corresponding unit vectors like e.g. $\hat{\boldsymbol{r}} \equiv \boldsymbol{r}/|\boldsymbol{r}|$ are dimensionless, since the dimensions cancel in the numerator and denominator.

We can also choose a basis consisting of two linearly independent dimensionless vectors $\boldsymbol{E}_1$ and $\boldsymbol{E}_2$, and write any vector in the plane as a linear combination of these two vectors, e.g.

$$\boldsymbol{r} = X^1\boldsymbol{E}_1 + X^2\boldsymbol{E}_2 = \sum_{i=1}^{2} X^i\boldsymbol{E}_i \equiv X^i\boldsymbol{E}_i \equiv \boldsymbol{E}_i X^i, \tag{2.1}$$

(where, of course, the components X^i have the dimension of a length). In the final steps of the previous equation I also introduced Einstein's *summation convention*: If in a multiplicative term an index appears exactly twice, then this index is automatically summed over its full range of possible values (in the example the index is i, and its possible values are 1 and 2). Writing the expansion coefficients behind the basis vectors, as in the last equation in (2.1), may look strange at first sight. However, there is certainly nothing wrong with it, and we will prefer this convention from now

doi:10.1088/2053-2571/aaf173ch2

on. The advantage of this convention is that transformation matrices will act on vector components X^i from the left, as we will see further below.

In the plane we would usually choose a Cartesian basis,

$$\boldsymbol{E}_i \cdot \boldsymbol{E}_j = \delta_{ij}, \tag{2.2}$$

but later we will find that in a general spacetime, Cartesian bases often can only be defined in limited regions of the spacetime. Therefore we consider the general case of an eventually non-Cartesian basis already now, i.e. we *do not assume* that the equations (2.2) holds.

If we cannot assume that our basis $\{\boldsymbol{E}_1, \boldsymbol{E}_2\}$ satisfies the equations (2.2), the length squared of the vector $\boldsymbol{r} = X^1\boldsymbol{E}_1 + X^2\boldsymbol{E}_2$ is

$$r^2 \equiv \boldsymbol{r}^2 = \boldsymbol{E}_i X^i \cdot \boldsymbol{E}_j X^j = \boldsymbol{E}_i \cdot \boldsymbol{E}_j X^i X^j = g_{ij} X^i X^j, \tag{2.3}$$

where we introduced the symbols

$$g_{ij} = \boldsymbol{E}_i \cdot \boldsymbol{E}_j. \tag{2.4}$$

Apparently these four symbols (g_{11}, $g_{12} = g_{21}$, g_{22}) define a symmetric matrix $\underline{g}$,

$$\underline{g} = \{g_{ij}\} = \begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix}. \tag{2.5}$$

We will use the designation **metric** for this group of real numbers, because they determine lengths. One might call $\underline{g}$ by the (awkward) name 'metric matrix'. Fortunately, we will see below that this matrix has special transformation properties under coordinate transformations in the plane, and matrices with these special transformation properties are denoted as 'tensors'. Therefore we end up with the much nicer designation of a *metric tensor* for $\underline{g}$.

The components of the inverse metric tensor $\underline{g}^{-1}$ are denoted by g^{ij}, and the equation $\underline{g}^{-1} \cdot \underline{g} = \underline{1}$ reads in components[1]

$$g^{ij} g_{jk} = g^i{}_k = \delta^i{}_k. \tag{2.7}$$

The *dual basis vectors* $\boldsymbol{E}^i$ are defined through the requirements

$$\boldsymbol{E}^i \cdot \boldsymbol{E}_k = \delta^i{}_k. \tag{2.8}$$

The dual basis vectors $\boldsymbol{E}^i$ must have an expansion in terms of the original basis vectors $\boldsymbol{E}_j$ since they are vectors in the same vector space, and we can easily

---

[1] At first sight, we should denote the components of $\underline{g}^{-1}$ as $(g^{-1})^{ij}$. However, we will see in a little while that we can pull indices up with $(g^{-1})^{ij}$, and if we pull up the indices of g_{ij} we find

$$g^{ij} \equiv (g^{-1})^{ik} g_{kl} (g^{-1})^{lj} = (g^{-1})^{ik} \delta_k{}^j = (g^{-1})^{ij}, \tag{2.6}$$

and therefore the notation $(g^{-1})^{ij}$ is redundant and never used. For the same reason we can write the unit matrix elements $\delta^i{}_k$ as $g^i{}_k$.

demonstrate that the components of the dual basis vectors with respect to the original basis are given by the components of the inverse metric, because

$$g^{ij}\boldsymbol{E}_j \cdot \boldsymbol{E}_k = g^{ij}g_{jk} = \delta^i{}_k, \tag{2.9}$$

i.e. the vector

$$\boldsymbol{E}^i = g^{ij}\boldsymbol{E}_j = g^{i1}\boldsymbol{E}_1 + g^{i2}\boldsymbol{E}_2 \tag{2.10}$$

satisfies exactly the defining property (2.8) of a dual basis vector.

The inversion of equation (2.10) is given by

$$\boldsymbol{E}_i = g_{ij}\boldsymbol{E}^j, \tag{2.11}$$

and we can derive that the inverse metric satisfies an equation similar to (2.4),

$$g^{ij} = g^{im}g_{mn}g^{nj} = g^{im}g^{jn}\boldsymbol{E}_m \cdot \boldsymbol{E}_n = \boldsymbol{E}^i \cdot \boldsymbol{E}^j. \tag{2.12}$$

For an example of a non-Cartesian basis and the related dual basis, consider the blue vectors $\boldsymbol{E}_1$ and $\boldsymbol{E}_2$ in figure 2.1. These vectors provide basis vectors $\boldsymbol{E}_i$ in the plane.

The angle between $\boldsymbol{E}_1$ and $\boldsymbol{E}_2$ is 60° or $\pi/3$ radian, and their lengths are $|\boldsymbol{E}_1| = 2$ and $|\boldsymbol{E}_2| = 1$. The metric tensor therefore has the following components in this basis,

$$\underline{g} = \begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix} = \begin{pmatrix} \boldsymbol{E}_1 \cdot \boldsymbol{E}_1 & \boldsymbol{E}_1 \cdot \boldsymbol{E}_2 \\ \boldsymbol{E}_2 \cdot \boldsymbol{E}_1 & \boldsymbol{E}_2 \cdot \boldsymbol{E}_2 \end{pmatrix} = \begin{pmatrix} 4 & 1 \\ 1 & 1 \end{pmatrix}. \tag{2.13}$$

The inverse matrix is then

$$\underline{g}^{-1} = \begin{pmatrix} g^{11} & g^{12} \\ g^{21} & g^{22} \end{pmatrix} = \frac{1}{3}\begin{pmatrix} 1 & -1 \\ -1 & 4 \end{pmatrix}. \tag{2.14}$$

This yields with equation (2.10) the dual basis vectors

$$\boldsymbol{E}^1 = \frac{1}{3}(\boldsymbol{E}_1 - \boldsymbol{E}_2), \quad \boldsymbol{E}^2 = \frac{1}{3}(4\boldsymbol{E}_2 - \boldsymbol{E}_1). \tag{2.15}$$

These equations determined the vectors $\boldsymbol{E}^i$ in figure 2.1.

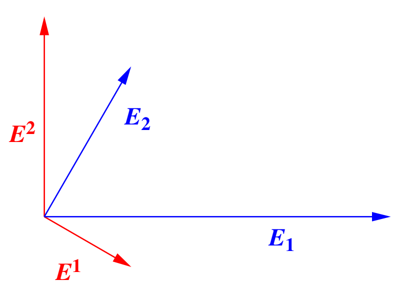

Figure 2.1. The blue set of vectors is the basis of vectors $\boldsymbol{E}_i$. The red vectors are the dual basis vectors $\boldsymbol{E}^i$.

Transformations

Now suppose that we would like to switch from our old set of basis vectors $\boldsymbol{E}_i$ to a new set of basis vectors $\boldsymbol{E}_i'$.

The new basis vectors are still part of the same vector space. Therefore there must exist an expansion of the new basis vectors in terms of the old basis vectors, i.e. there must exist a set of equations of the form

$$\boldsymbol{E}_i' = \boldsymbol{E}_j M^j{}_i. \tag{2.16}$$

How would the components X'^i of a vector $\boldsymbol{r}$ with respect to the new basis vectors relate to the old components X^j? The expansion of the same vector $\boldsymbol{r}$ in both bases yields $\boldsymbol{r} = \boldsymbol{E}_i' X'^i = \boldsymbol{E}_j X^j$. Substitution of equation (2.16) yields $\boldsymbol{r} = \boldsymbol{E}_j M^j{}_i X'^i = \boldsymbol{E}_j X^j$, and since the components of a $\boldsymbol{r}$ with respect to the basis vectors $\boldsymbol{E}_j$ are unique, we must have $M^j{}_i X'^i = X^j$, or

$$X'^i = (M^{-1})^i{}_j X^j = X^j (M^{-1,\,T})_j{}^i, \tag{2.17}$$

i.e. if the vectors $\boldsymbol{E}_i$ transform with the matrix $\underline{M}$ from the right (2.16), then the corresponding coordinates X^i transform with the *inverse* matrix $\underline{M}^{-1}$ from the left (2.17) (or with the *contragredient matrix* $\underline{M}^{-1,\,T}$ from the right). The coordinates seem to transform in a sense 'opposite' to the basis vectors, and the mathematical expression for 'opposite' is *contravariant*. The transformation law (2.16) for the basis vectors is denoted as *covariant*. We can also calculate how the new and the old components of the metric are related: We have

$$g_{ij}' = \boldsymbol{E}_i' \cdot \boldsymbol{E}_j' = \boldsymbol{E}_m M^m{}_i \cdot \boldsymbol{E}_n M^n{}_j = g_{mn} M^m{}_i M^n{}_j, \tag{2.18}$$

i.e. the components of the metric transform with the same transformation coefficients $M^i{}_j$ as the basis vectors, only that there are two copies of them appearing in the transformation law. For the inverse metric one finds the transformation law

$$g'^{ij} = (M^{-1})^i{}_m (M^{-1})^j{}_n \, g^{mn}. \tag{2.19}$$

A geometrical object whose components transform with a combination of coefficients $M^i{}_j$ or $(M^{-1})^i{}_j$ is denoted as a *tensor*. The tensor is denoted as a *covariant tensor of order n* if n coefficients $M^i{}_j$ appear in the transformation law. The tensor is denoted as a *contravariant tensor of order n* if n copies of coefficients $(M^{-1})^i{}_j$ appear in the transformation law. E.g. the metric is a covariant tensor of second order, whereas the inverse metric is a contravariant tensor of second order.

We only transformed the basis vectors $\boldsymbol{E}_i$ of our space (here: the two-dimensional plane), i.e. we only changed the basis vectors but did nothing to our actual space: $\boldsymbol{r}$ is still one and the same vector, it is only decomposed in two different bases. Such a transformation which does not do anything to our actual space is denoted as a *passive transformation*, and this is what we mainly have to deal with in general relativity. A transformation where we actually rotate or stretch or in any way deform our two-dimensional plane would be an example of an *active transformation*.

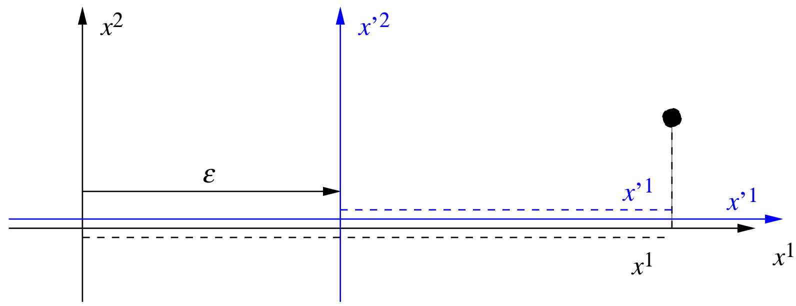

Figure 2.2. A translation of the *coordinate system* by ϵ^1 changes the x^1 coordinate to $x'^1 = x^1 - \epsilon^1$. This is the *passive* interpretation of a transformation.

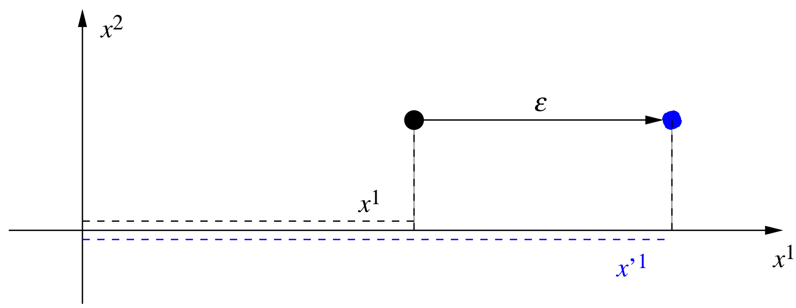

Figure 2.3. A translation of the *space* by ϵ^1 in the x^1 direction changes the x^1 coordinate to $x'^1 = x^1 + \epsilon^1$. This is the *active* interpretation of a transformation.

Figures 2.2 and 2.3 explain the difference between passive and active transformations in the case of translations.

For an *active* linear transformation, the equation

$$\boldsymbol{r}' = \boldsymbol{E}'_j X^j = \boldsymbol{E}_i M^i{}_j X^j \equiv \boldsymbol{E}_i X'^i \tag{2.20}$$

implies for the coordinates the transformation law

$$X'^i = M^i{}_j X^j. \tag{2.21}$$

Covariant and contravariant vector components

We have seen that for a given set of basis vectors $\{\boldsymbol{E}_i\}$ there is a second set of basis vectors $\{\boldsymbol{E}^i\}$ provided by the dual basis vectors. We can decompose a vector $\boldsymbol{r}$ in whichever basis we like, i.e. instead of the decomposition $\boldsymbol{r} = \boldsymbol{E}_i X^i$ we can also use the decomposition with respect to the dual basis, $\boldsymbol{r} = \boldsymbol{E}^i X_i$. Substitution of equation (2.10) then yields

$$\boldsymbol{r} = \boldsymbol{E}^i X_i = \boldsymbol{E}_j g^{ij} X_i, \tag{2.22}$$

i.e. the vector components with respect to the two dual bases are related by (remember the symmetry of the metric)

$$X^j = X_i g^{ij} = g^{ji} X_i, \quad X_i = X^j g_{ji}. \tag{2.23}$$

The contravariant transformation property of X^j and the covariant transformation property of g_{ij} implies that the dual vector components X_i transform covariantly under passive transformations (2.16), (2.17),

$$X_i' = X'^j g'_{ji} = (M^{-1})^j{}_c X^c g_{ab} M^a{}_j M^b{}_i = X^a g_{ab} M^b{}_i = X_b M^b{}_i. \tag{2.24}$$

This is a general recipe: every lower index on a tensor transforms covariantly and every upper index transforms contravariantly under passive transformations.

Scalar products revisited

The scalar product of two vectors $\boldsymbol{u} = \boldsymbol{E}_i u^i$ and $\boldsymbol{v} = \boldsymbol{E}_i v^i$ is with the previous definitions of metric tensor and covariant coordinates,

$$\boldsymbol{u} \cdot \boldsymbol{v} = \boldsymbol{E}_i \cdot \boldsymbol{E}_j u^i v^j = g_{ij} u^i v^j = u^i v_i = u_j v^j. \tag{2.25}$$

The opposite transformation properties of covariant and contravariant indices implies invariance of the scalar product under transformations:

$$u'_i v'^i = u_a M^a{}_i (M^{-1})^i{}_b v^b = u_a v^a. \tag{2.26}$$

This holds in general for summation over a covariant and a contravariant index.

Tensor products

The scalar product of two vectors produces a number. However, we can also multiply two vectors $\boldsymbol{u} = \boldsymbol{E}_i u^i$ and $\boldsymbol{v} = \boldsymbol{E}_i v^i$ to generate a tensor, denoted as $\boldsymbol{u} \otimes \boldsymbol{v}^T$. For the start, we can think of the tensor $\boldsymbol{u} \otimes \boldsymbol{v}^T$ as the matrix with components $u^i v_j$,

$$(\boldsymbol{u} \otimes \boldsymbol{v}^T)^i{}_j = u^i v_j = u^i v^k g_{kj}. \tag{2.27}$$

In fact, we will also use the matrices with components

$$(\boldsymbol{u} \otimes \boldsymbol{v}^T)^{ij} = u^i v^j, \quad (\boldsymbol{u} \otimes \boldsymbol{v}^T)_i{}^j = u_i v^j = g_{ik} u^k v^j, \tag{2.28}$$

$$(\boldsymbol{u} \otimes \boldsymbol{v}^T)_{ij} = u_i v_j = g_{im} g_{jn} u^m v^n. \tag{2.29}$$

Since we can pull indices up and down as we please using the metric tensor, all these matrices are easily transformed into each other and have all the same information content. Mathematically we should think of them as different representations of one and the same mathematical object, the tensor $\boldsymbol{u} \otimes \boldsymbol{v}^T$.

We often simply write $\boldsymbol{u} \otimes \boldsymbol{v}$ instead of $\boldsymbol{u} \otimes \boldsymbol{v}^T$, just like we simply write $\boldsymbol{u} \cdot \boldsymbol{v}$ instead of $\boldsymbol{u}^T \cdot \boldsymbol{v}$ for scalar products.

Tensor products appear automatically in projections of vectors. Suppose two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$. We want to decompose the vector $\boldsymbol{b}$ into two vectors $\boldsymbol{b}_{\|}$ parallel to $\boldsymbol{a}$ and $\boldsymbol{b}_{\perp}$ perpendicular to $\boldsymbol{a}$ (figure 2.4).

We need the unit vector $\hat{\boldsymbol{a}} = \boldsymbol{a}/|\boldsymbol{a}|$ for the construction. We know $\boldsymbol{b}_{\|}\|\hat{\boldsymbol{a}}$,

$$\boldsymbol{b}_{\|} = |\boldsymbol{b}_{\|}|\hat{\boldsymbol{a}}. \tag{2.30}$$

To figure out $|\boldsymbol{b}_{\|}|$, we observe that $|\boldsymbol{b}_{\|}| = |\boldsymbol{b}|\cos\alpha = |\hat{\boldsymbol{a}}||\boldsymbol{b}|\cos\alpha = \hat{\boldsymbol{a}} \cdot \boldsymbol{b}$, i.e. we can write

$$\boldsymbol{b}_{\|} = \hat{\boldsymbol{a}}(\hat{\boldsymbol{a}} \cdot \boldsymbol{b}). \tag{2.31}$$

Let us repeat this equation in terms of vector components with respect to some basis $\boldsymbol{E}_i$. Substitution of $\boldsymbol{b}_{\|} = b_{\|}^i \boldsymbol{E}_i$, $b = b^i \boldsymbol{E}_i$, and $\hat{\boldsymbol{a}} = \hat{a}^i \boldsymbol{E}_i$ yields

$$\boldsymbol{b}_{\|} = \boldsymbol{E}_i b_{\|}^i = \boldsymbol{E}_i \hat{a}^i \hat{a}^k \boldsymbol{E}_k \cdot \boldsymbol{E}_j b^j = \boldsymbol{E}_i \hat{a}^i \hat{a}^k b^j g_{kj} = \boldsymbol{E}_i \hat{a}^i \hat{a}_j b^j, \tag{2.32}$$

or

$$b_{\|}^i = \hat{a}^i \hat{a}_j b^j. \tag{2.33}$$

We can think of this as a matrix multiplication

$$\boldsymbol{b}_{\|} = (\hat{\boldsymbol{a}} \otimes \hat{\boldsymbol{a}}^T) \cdot \boldsymbol{b}, \tag{2.34}$$

where the matrix $\hat{\boldsymbol{a}} \otimes \hat{\boldsymbol{a}}^T$ has components

$$(\hat{\boldsymbol{a}} \otimes \hat{\boldsymbol{a}}^T)^i{}_j = \hat{a}^i \hat{a}_j. \tag{2.35}$$

From equation (2.34) we find for the perpendicular vector

$$\boldsymbol{b}_{\perp} = \boldsymbol{b} - \boldsymbol{b}_{\|} = (\underline{1} - \hat{\boldsymbol{a}} \otimes \hat{\boldsymbol{a}}^T) \cdot \boldsymbol{b}. \tag{2.36}$$

The message from equations (2.34), (2.36) is that multiplication of any vector $\boldsymbol{b}$ with the *projector*

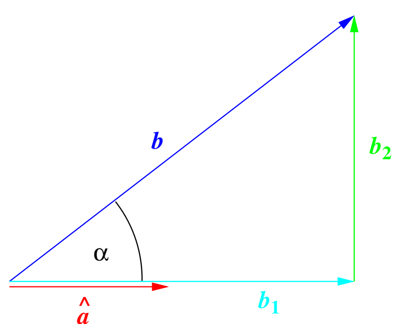

Figure 2.4. The cyan vector is the projection of the blue vector $\boldsymbol{b}$ onto the red unit vector $\hat{\boldsymbol{a}}$, $\boldsymbol{b}_1 = \boldsymbol{b}_{\|}$. The green vector is the projection of the vector $\boldsymbol{b}$ orthogonal to the unit vector $\hat{\boldsymbol{a}}$, $\boldsymbol{b}_2 = \boldsymbol{b}_{\perp}$.

$$\underline{P}_{\parallel} = \hat{\boldsymbol{a}} \otimes \hat{\boldsymbol{a}}^T \tag{2.37}$$

projects the vector onto its component parallel to the unit vector $\hat{\boldsymbol{a}}$, and multiplication with the perpendicular projector

$$\underline{P}_{\perp} = \underline{1} - \hat{\boldsymbol{a}} \otimes \hat{\boldsymbol{a}}^T = \underline{1} - \underline{P}_{\parallel} \tag{2.38}$$

projects any vector onto its component orthogonal to $\hat{\boldsymbol{a}}$.

Higher-order tensors and tensor products

The transformation properties of basis vectors and vector components

$$\boldsymbol{E}_i \to \boldsymbol{E}_i' = \boldsymbol{E}_j M^j{}_i, \quad X^i \to X'^i = (M^{-1})^i{}_j X^j, \tag{2.39}$$

imply that vectors are invariant under transformations of the basis,

$$\boldsymbol{r} = \boldsymbol{E}_i X^i = \boldsymbol{E}_i' X'^i. \tag{2.40}$$

Indeed, we have determined the contravariant transformation behavior of the vector components from the requirement that the vector $\boldsymbol{r}$ itself can be expanded in either basis, but does not depend on which basis we choose. However, then we also found matrices T^{ij} (e.g. g^{ij}) with the transformation property

$$T^{ij} \to T'^{ij} = (M^{-1})^i{}_a (M^{-1})^j{}_b T^{ab}. \tag{2.41}$$

This implies that the following combination of matrix components and basis vectors is also invariant under the transformation (2.16):

$$\underline{T} = \boldsymbol{E}_i \otimes \boldsymbol{E}_j T^{ij} = \boldsymbol{E}_i' \otimes \boldsymbol{E}'{}_j T'^{ij}. \tag{2.42}$$

The symbol $\boldsymbol{E}_i \otimes \boldsymbol{E}_j$ indicates that this is just a bilinear combination of the two basis vectors $\boldsymbol{E}_i$ and $\boldsymbol{E}_j$, but no scalar product between the two vectors is involved. We can think of the *tensor* $\underline{T}$ as a higher-order generalization of a vector. A vector is a linear combination of basis vectors,

$$\boldsymbol{r} = \boldsymbol{E}_i X^i = X_i \boldsymbol{E}^i, \quad X_i = g_{ij} X^j, \tag{2.43}$$

and a tensor is a linear combination of 'basis dyads',

$$\underline{T} = \boldsymbol{E}_i \otimes \boldsymbol{E}_j T^{ij} = \boldsymbol{E}_i \otimes \boldsymbol{E}^j T^i{}_j = \boldsymbol{E}^i \otimes \boldsymbol{E}_j T_i{}^j = \boldsymbol{E}^i \otimes \boldsymbol{E}^j T_{ij}, \tag{2.44}$$

$$T^i{}_j = g_{jk} T^{ik}, \quad T_i{}^j = g_{ik} T^{kj}, \quad T_{ij} = g_{ik} g_{jl} T^{kl}. \tag{2.45}$$

The tensor product or dyadic product of two vectors is then simply

$$\boldsymbol{u} \otimes \boldsymbol{v}^T = \boldsymbol{E}_i \otimes \boldsymbol{E}_j u^i v^j = \boldsymbol{E}_i \otimes \boldsymbol{E}^j u^i v_j = \boldsymbol{E}^i \otimes \boldsymbol{E}_j u_i v^j = \boldsymbol{E}^i \otimes \boldsymbol{E}^j u_i v_j. \tag{2.46}$$

Once this level of abstraction is attained, it is trivial to go to higher order, e.g.

$$\underline{V} = \boldsymbol{E}_i \otimes \boldsymbol{E}_j \otimes \boldsymbol{E}_k V^{ijk} = \boldsymbol{E}_i \otimes \boldsymbol{E}_j \otimes \boldsymbol{E}^k V^{ij}{}_k = \dots \tag{2.47}$$

with

$$V'^{ijk} = (M^{-1})^i{}_a (M^{-1})^j{}_b (M^{-1})^k{}_c V^{abc} \tag{2.48}$$

under basis transformations (2.16). The tensor $\underline{V}$ is a third-order tensor, while $\underline{T}$ is a second-order tensor and $\boldsymbol{r}$ is a first-order tensor. We can also form higher-order tensor products, where the product of a tensor $\underline{U}$ of order m with a tensor $\underline{V}$ of order n yields a tensor of order $m + n$, in components

$$(\underline{U} \otimes \underline{V})^{i_1 \dots i_{m+n}} = U^{i_1 \dots i_m} V^{i_1 \dots i_n}. \tag{2.49}$$

The formulation of special relativity requires second-order tensors, and we will encounter fourth-order tensors in general relativity.

2.2 Curvilinear coordinates

Now we will do something even more complicated and introduce more general coordinate transformations than the linear transformations (2.17).

Suppose that X^1 and X^2 are Cartesian coordinates in the plane, and $\boldsymbol{e}_1 \equiv \boldsymbol{e}_x$, $\boldsymbol{e}_2 \equiv \boldsymbol{e}_y$ are the corresponding orthonormal basis vectors,

$$\boldsymbol{e}_i \cdot \boldsymbol{e}_j = \delta_{ij}. \tag{2.50}$$

Then we write a transformation to new coordinates x^1 and x^2 as

$$X^1, X^2 \to x^1(X^1, X^2), x^2(X^1, X^2), \tag{2.51}$$

or in shorthand $X^i \to x^\mu(X)$, and the inverse transformation is in shorthand $x^\mu \to X^i(x)$. This defines *bona fide* coordinate transformations in those parts of the plane where the relations $X^i \to x^\mu(X)$, $x^\mu \to X^i(x)$ are invertible (and usually we also must have differentiability to some appropriate order).

Usually changing one of the coordinates x^μ to $x^\mu + \Delta x^\mu$ will not take us along straight lines any more, i.e. the x^μ coordinate lines will generically be curved.

A vector $\boldsymbol{r}$ in the Cartesian basis is

$$\boldsymbol{r}(x) = X^i(x)\boldsymbol{e}_i, \tag{2.52}$$

and the vector $d\boldsymbol{r}(x)$ connecting the two points $x = (x^1, x^2)$ and $x + dx = (x^1 + dx^1, x^2 + dx^2)$ is

$$d\boldsymbol{r}(x) = \boldsymbol{r}(x + dx) - \boldsymbol{r}(x) = dx^\mu \partial_\mu \boldsymbol{r}(x) = dx^\mu \boldsymbol{E}_\mu(x). \tag{2.53}$$

This tells us that

$$\boldsymbol{E}_\mu(x) = \partial_\mu \boldsymbol{r} = \partial_\mu X^i(x)\boldsymbol{e}_i, \tag{2.54}$$

is a tangent vector to the generically curved x^μ-coordinate line, and its Cartesian components are given by the Jacobian matrix of the coordinate transformation,

$$E_\mu{}^i(x) = \partial_\mu X^i(x). \tag{2.55}$$

Equation (2.53) yields for the distance squared between the two points $x = (x^1, x^2)$ and $x + dx = (x^1 + dx^1, x^2 + dx^2)$ the result

$$ds^2 = d\boldsymbol{r}^2(x) = dx^\mu dx^\nu \boldsymbol{E}_\mu(x) \cdot \boldsymbol{E}_\nu(x) = g_{\mu\nu}(x) dx^\mu dx^\nu, \tag{2.56}$$

with the components of the metric tensor in the new coordinates:

$$g_{\mu\nu}(x) = \boldsymbol{E}_\mu(x) \cdot \boldsymbol{E}_\nu(x) = \partial_\mu X^i(x)\boldsymbol{e}_i \cdot \partial_\nu X^j(x)\boldsymbol{e}_j = \partial_\mu X^i(x) \cdot \partial_\nu X^j(x)\delta_{ij}. \tag{2.57}$$

The product of the inverse Jacobian matrices

$$g^{\mu\nu}(x) = \delta^{ij}\partial_i x^\mu \cdot \partial_j x^\nu \tag{2.58}$$

then yields the inverse metric:

$$\begin{aligned} g^{\mu\rho}(x) g_{\rho\nu}(x) &= \delta^{ij}\partial_i x^\mu \cdot \partial_j x^\rho \cdot \partial_\rho X^k \cdot \partial_\nu X^l \delta_{kl} = \delta^{ij}\partial_i x^\mu \cdot \partial_j X^k \cdot \partial_\nu X^l \delta_{kl} \\ &= \delta^{ij}\partial_i x^\mu \cdot \delta_j{}^k \cdot \partial_\nu X^l \delta_{kl} = \delta^{ij}\partial_i x^\mu \cdot \partial_\nu X^l \delta_{jl} = \partial_\nu X^i \cdot \partial_i x^\mu \\ &= \partial_\nu x^\mu = \delta^\mu{}_\nu. \end{aligned} \tag{2.59}$$

The components $g^{\mu\nu}(x)$ of the inverse metric tensor yield again dual vectors in the point x:

$$\boldsymbol{E}^\mu(x) = g^{\mu\nu}(x)\boldsymbol{E}_\nu(x), \tag{2.60}$$

with scalar products

$$\boldsymbol{E}^\mu(x) \cdot \boldsymbol{E}_\kappa(x) = g^{\mu\nu}(x) g_{\nu\kappa}(x) = \delta^\mu{}_\kappa, \quad \boldsymbol{E}^\mu(x) \cdot \boldsymbol{E}^\nu(x) = g^{\mu\nu}(x). \tag{2.61}$$

Substitution of equation (2.58) into (2.60) also yields the following representation of the dual basis vectors,

$$\boldsymbol{E}^\mu(x) = \delta^{ij}\partial_i x^\mu \cdot \partial_j x^\nu \cdot \partial_\nu X^k \boldsymbol{e}_k = \delta^{ij}\partial_i x^\mu \delta_j{}^k \boldsymbol{e}_k = \boldsymbol{e}^i \partial_i x^\mu. \tag{2.62}$$

We will also often use normalized tangent vectors besides the induced tangent vectors $\boldsymbol{E}_\mu(x)$ (NO summation convention in the following equation!)

$$\boldsymbol{e}_\mu(x) = \frac{\boldsymbol{E}_\mu(x)}{|\boldsymbol{E}_\mu(x)|} = \frac{\boldsymbol{E}_\mu(x)}{\sqrt{g_{\mu\mu}(x)}}, \quad \boldsymbol{e}_\mu(x) \cdot \boldsymbol{e}_\nu(x) = \frac{g_{\mu\nu}(x)}{\sqrt{g_{\mu\mu}(x) g_{\nu\nu}(x)}}. \tag{2.63}$$

Calculation of the length of a line

Suppose we have a (generically curved) line in the plane which takes us from the point x_1 to the point x_2. We can parametrize this line with an arbitrary monotonically increasing real parameter τ: $x = x(\tau)$, e.g. such that $x_1 = x(0)$ and $x_2 = x(1)$ (or in coordinates: $x_1^\mu = x^\mu(0)$ and $x_2^\mu = x^\mu(1)$). Equation (2.56) tells us that the length element along the line is

$$ds(\tau) = \sqrt{g_{\mu\nu}(x)dx^\mu(\tau)dx^\nu(\tau)} = d\tau\sqrt{g_{\mu\nu}(x)\frac{dx^\mu}{d\tau}\frac{dx^\nu}{d\tau}}, \tag{2.64}$$

and we can calculate the whole length of the line through integration of the line element along the whole line:

$$s = \int ds = \int_0^1 d\tau\sqrt{g_{\mu\nu}(x)\frac{dx^\mu}{d\tau}\frac{dx^\nu}{d\tau}} = \int_0^1 d\tau\sqrt{g_{\mu\nu}(x)\dot{x}^\mu\dot{x}^\nu}. \tag{2.65}$$

∇ in general coordinates

We can derive the gradient or del operator in general curvilinear coordinates from its definition in a Cartesian basis $\boldsymbol{e}^i(=\boldsymbol{e}_i)$,

$$\nabla = \boldsymbol{e}^i\partial_i = \boldsymbol{e}^i\partial_i x^\mu\partial_\mu = \boldsymbol{E}^\mu(x)\partial_\mu, \tag{2.66}$$

where we used the result (2.62) for the dual basis vectors in general curvilinear coordinates. The representation of the gradient operator in a general coordinate system is therefore

$$\nabla = \boldsymbol{E}^\mu(x)\partial_\mu. \tag{2.67}$$

We can see these notions at work in polar coordinates.

Polar coordinates in the plane

Here we use $X^1 \equiv X$, $X^2 \equiv Y$ for the Cartesian coordinates in the plane. The curvilinear coordinates are $x^1 = r$ and $x^2 = \varphi$. The Cartesian basis vectors are $\boldsymbol{e}_x$, $\boldsymbol{e}_y$. The transformation laws between the Cartesian and polar coordinates are

$$r = \sqrt{X^2 + Y^2}, \quad \varphi = \arctan(Y/X), \quad X = r\cos\varphi, \quad Y = r\sin\varphi, \tag{2.68}$$

and the tangent vectors along the (r, φ) coordinate lines are

$$\boldsymbol{E}_r = \partial_r\boldsymbol{r} = \boldsymbol{e}_x\cos\varphi + \boldsymbol{e}_y\sin\varphi, \tag{2.69}$$

$$\boldsymbol{E}_\varphi = \partial_\varphi\boldsymbol{r} = -\boldsymbol{e}_x r\sin\varphi + \boldsymbol{e}_y r\cos\varphi. \tag{2.70}$$

The components of the metric tensor in polar coordinates are therefore

$$\{g_{\mu\nu}\} = \begin{pmatrix} g_{rr} & g_{r\varphi} \\ g_{\varphi r} & g_{\varphi\varphi} \end{pmatrix} = \begin{pmatrix} \boldsymbol{E}_r\cdot\boldsymbol{E}_r & \boldsymbol{E}_r\cdot\boldsymbol{E}_\varphi \\ \boldsymbol{E}_\varphi\cdot\boldsymbol{E}_r & \boldsymbol{E}_\varphi\cdot\boldsymbol{E}_\varphi \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & r^2 \end{pmatrix}, \tag{2.71}$$

$$\{g^{\mu\nu}\} = \begin{pmatrix} 1 & 0 \\ 0 & r^{-2} \end{pmatrix}, \tag{2.72}$$

and the length element (2.56) is $ds^2 = dr^2 + r^2 d\varphi^2$.

The dual tangent vectors are

$$\boldsymbol{E}^r = \boldsymbol{E}_r \equiv \boldsymbol{e}_r, \quad \boldsymbol{E}^\varphi = \frac{1}{r^2}\boldsymbol{E}_\varphi = \frac{1}{r}\boldsymbol{e}_\varphi = -\boldsymbol{e}_x\frac{\sin\varphi}{r} + \boldsymbol{e}_y\frac{\cos\varphi}{r}, \tag{2.73}$$

and the scaling and orientation of the basis vectors and dual basis vectors are illustrated in figure 2.5.

The gradient operator (2.67) in polar coordinates is

$$\nabla = \boldsymbol{E}^r\frac{\partial}{\partial r} + \boldsymbol{E}^\varphi\frac{\partial}{\partial\varphi} = \boldsymbol{e}_r\frac{\partial}{\partial r} + \frac{1}{r}\boldsymbol{e}_\varphi\frac{\partial}{\partial\varphi}. \tag{2.74}$$

From this follows directly the Laplace operator in polar coordinates

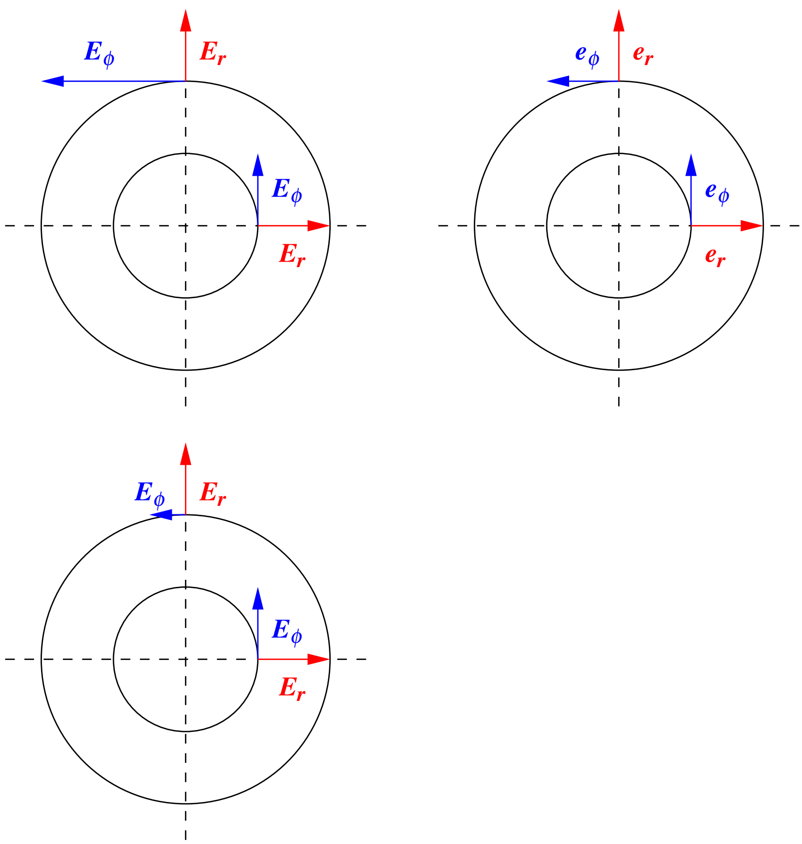

Figure 2.5. The scaling of the various tangent vector bases for polar coordinates with distance.

$$\begin{aligned}\Delta = \nabla^2 = \nabla \cdot \nabla &= \frac{\partial^2}{\partial r^2} + \frac{1}{r}\boldsymbol{e}_\varphi \cdot \frac{\partial \boldsymbol{e}_r}{\partial \varphi}\frac{\partial}{\partial r} + \frac{1}{r^2}\frac{\partial^2}{\partial \varphi^2} \\ &= \frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r} + \frac{1}{r^2}\frac{\partial^2}{\partial \varphi^2}.\end{aligned} \tag{2.75}$$

The Laplace operator in general coordinates

The definition (2.67) yields for a function $f(x)$

$$\begin{aligned}\Delta f(x) = \nabla^2 f(x) &= \boldsymbol{E}^\mu(x)\partial_\mu \cdot \boldsymbol{E}^\nu(x)\partial_\nu f(x) \\ &= g^{\mu\nu}(x)\partial_\mu\partial_\nu f(x) + \boldsymbol{E}^\mu(x) \cdot (\partial_\mu \boldsymbol{E}^\nu(x))\partial_\nu f(x) \\ &= g^{\mu\nu}(x)\big(\partial_\mu\partial_\nu f(x) + \boldsymbol{E}_\nu(x) \cdot (\partial_\mu \boldsymbol{E}^\rho(x))\partial_\rho f(x)\big).\end{aligned} \tag{2.76}$$

In the next step we use

$$\boldsymbol{E}_\nu(x) \cdot \boldsymbol{E}^\rho(x) = \delta_\nu{}^\rho \quad \Rightarrow \quad \boldsymbol{E}_\nu(x) \cdot \partial_\mu \boldsymbol{E}^\rho(x) = -\boldsymbol{E}^\rho(x) \cdot \partial_\mu \boldsymbol{E}_\nu(x) \tag{2.77}$$

to find

$$\begin{aligned}\Delta f(x) &= g^{\mu\nu}(x)\big(\partial_\mu\partial_\nu f(x) - \boldsymbol{E}^\rho(x) \cdot (\partial_\mu \boldsymbol{E}_\nu(x))\partial_\rho f(x)\big) \\ &= g^{\mu\nu}(x)\big(\partial_\mu\partial_\nu f(x) - \Gamma^\rho{}_{\nu\mu}(x)\partial_\rho f(x)\big),\end{aligned} \tag{2.78}$$

with the *connection coefficients* or *Christoffel symbols*

$$\Gamma^\rho{}_{\nu\mu}(x) = \boldsymbol{E}^\rho(x) \cdot \partial_\mu \boldsymbol{E}_\nu(x). \tag{2.79}$$

We can express the Christoffel symbols through the metric,

$$\begin{aligned}\Gamma^\rho{}_{\nu\mu} &= \boldsymbol{E}^\rho \cdot \partial_\mu \boldsymbol{E}_\nu = \boldsymbol{E}^\rho \cdot \partial_\mu\partial_\nu \boldsymbol{r} = \boldsymbol{E}^\rho \cdot \partial_\nu \boldsymbol{E}_\mu \\ &= \frac{1}{2}g^{\rho\sigma}\big(\boldsymbol{E}_\sigma \cdot \partial_\mu \boldsymbol{E}_\nu + \boldsymbol{E}_\sigma \cdot \partial_\nu \boldsymbol{E}_\mu\big) \\ &= \frac{1}{2}g^{\rho\sigma}\big[\partial_\mu(\boldsymbol{E}_\sigma \cdot \boldsymbol{E}_\nu) + \partial_\nu\big(\boldsymbol{E}_\sigma \cdot \boldsymbol{E}_\mu\big) - \partial_\sigma\big(\boldsymbol{E}_\nu \cdot \boldsymbol{E}_\mu\big)\big] \\ &= \frac{1}{2}g^{\rho\sigma}\big(\partial_\mu g_{\sigma\nu} + \partial_\nu g_{\sigma\mu} - \partial_\sigma g_{\nu\mu}\big).\end{aligned} \tag{2.80}$$

The only non-vanishing Christoffel symbols for polar coordinates in the plane are with $x^1 = r$, $x^2 = \varphi$,

$$\Gamma^1{}_{22} = \Gamma^r{}_{\varphi\varphi} = -r, \quad \Gamma^2{}_{12} = \Gamma^\varphi{}_{r\varphi} = \Gamma^\varphi{}_{\varphi r} = \frac{1}{r}, \tag{2.81}$$

where the convention to use the coordinates themselves as indices is also used. This is a common practice in particular for polar coordinate systems.

Chapter 3

The tangents of spacetime: special relativity

The inception of special relativity preceded the inception of general relativity, and the latter theory generalized special relativity to solve the shortcoming that special relativity could not accommodate gravity. However, special relativity is still the theory of choice for the description of fast motion or high-energy phenomena which are dominated by non-gravitational interactions. We will recognize in chapter 7 that the requirement of negligibility of gravity is equivalent to the requirement of the negligibility of the local curvature of spacetime over the length and time scales of an experiment. Stated differently, special relativity is applicable when we can approximate the four-dimensional curved spacetime through its flat local four-dimensional tangent spacetime, akin to neglecting the local curvature of the Earth's surface in observations over small length scales. In that sense, special relativity naturally lives in the flat four-dimensional tangent planes of spacetime.

3.1 Lorentz transformations and the relativity of space and time

We need an operational definition for the notion of inertial frames, i.e. coordinate systems where the laws of nature take a particularly simple form. It has already been mentioned in section 1.3 that reference to the cosmic rest frame or reference to free particles does not provide satisfactory definitions for inertial frames, but we can use the following definition: An inertial frame is a coordinate system $\{t, \boldsymbol{x}\}$ where light fronts in vacuum move according to the equation $c^2\Delta t^2 = \Delta \boldsymbol{x}^2$.

The easiest approach to special relativity uses the outcome of the Michelson experiment as an indication that two uniformly moving observers will always measure the same value $c = 299\,792\,458$ m s^{-1} for the vacuum speed of light in every direction, irrespective of their constant relative velocity $\boldsymbol{v}$, i.e. if Frank uses his reference frame to measure velocity $\boldsymbol{c}$ for a light wave, Mary will also measure velocity $\boldsymbol{c}$ in her reference frame, but *not* the velocity $\boldsymbol{c} - \boldsymbol{v}$ which would be predicted by the Galilei boost (1.4). The Galilei boost (1.4) must therefore be replaced with a

doi:10.1088/2053-2571/aaf173ch3

coordinate transformation $\{t, \boldsymbol{x}\} \to \{t', \boldsymbol{x}'\}$ with the property of leaving the vacuum speed of light invariant in every direction,

$$c^2\Delta t^2 = \Delta \boldsymbol{x}^2 \quad \Leftrightarrow \quad c^2\Delta t'^2 = \Delta \boldsymbol{x}'^2. \tag{3.1}$$

This implies that the Galilei boost (1.4) between the coordinates of Frank and Mary must be replaced by the *Lorentz boost*

$$ct' = \frac{ct - (\boldsymbol{v}/c)\cdot \boldsymbol{x}}{\sqrt{1 - (v^2/c^2)}}, \tag{3.2}$$

$$\boldsymbol{x}' = \boldsymbol{x} + \left(\frac{1}{v^2\sqrt{1 - (v^2/c^2)}} - \frac{1}{v^2}\right)(\boldsymbol{v}\cdot \boldsymbol{x})\boldsymbol{v} - \frac{\boldsymbol{v}t}{\sqrt{1 - (v^2/c^2)}}. \tag{3.3}$$

This looks rather cumbersome, and we will find much neater and more useful ways to present this result in section 3.3. However, for now it is sufficient to simply accept the result and specialize to the case of constant relative velocity in the x-direction, $\boldsymbol{v} = v\boldsymbol{e}_x$, when the transformation laws (3.2) and (3.3) reduce to

$$ct' = \frac{ct - (v/c)x}{\sqrt{1 - (v^2/c^2)}}, \quad x' = \frac{x - vt}{\sqrt{1 - (v^2/c^2)}}, \quad y' = y, \quad z' = z. \tag{3.4}$$

To discuss the consequences of these transformation laws, it is convenient to introduce $\beta = v/c$ and $\gamma = (1 - \beta^2)^{-1/2}$. The transformation equations (3.4) and their inversions then read

$$ct' = \gamma(ct - \beta x), \quad x' = \gamma(x - \beta ct), \quad y' = y, \quad z' = z, \tag{3.5}$$

$$ct = \gamma(ct' + \beta x'), \quad x = \gamma(x' + \beta ct'). \tag{3.6}$$

Mary is located in the spatial origin $x' = y' = z' = 0$ of her coordinate system (ct', x', y', z'). Substitution of $x' = 0$ into the second equation in (3.5) then tells us that she is moving along the line $x = \beta ct = vt$, $y = z = 0$ in Frank's coordinate system (ct, x, y, z), i.e. she moves with velocity $\boldsymbol{v} = v\boldsymbol{e}_x$ relative to Frank. On the other hand, Frank's coordinates $x = y = z = 0$ imply $x' = -\beta ct'$, i.e. the spatial origin of Frank's system moves with velocity $-v\boldsymbol{e}'_x$ relative to Mary's system. Equations (3.5) and (3.6) therefore describe the Lorentz boost between two coordinate frames with a relative motion with speed v in the x-direction (as seen in Frank's frame), or in the $(-x')$-direction (as seen in Mary's frame), with coinciding origins at times $t = t' = 0$, see figure 3.1.

Note that the three-dimensional spatial slice $t = 0$ that Frank might use to define a spatial section of the four-dimensional Universe differs from Mary's spatial section $t' = 0$. This is the relativity of space and time. Time is not universal for different observers, and therefore the assignment of time values to events, as well as the definition of 'the spatial Universe now', are different for different observers. In particular, figure 3.1 depicts three-dimensional sections of the two four-dimensional

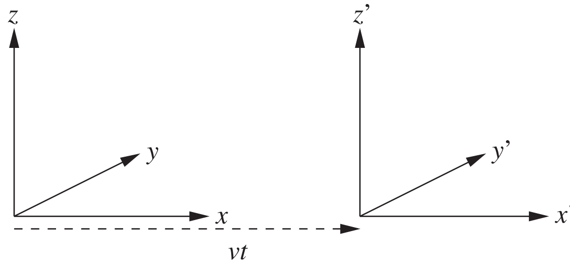

Figure 3.1. Two frames with relative velocity $\boldsymbol{v} = v\boldsymbol{e}_x$.

reference frames of Frank and Mary at some value of Frank's time t, because we denoted the distance between Frank and Mary as vt. Mary's picture at time t' would look the same, with Frank being a distance vt' away. However, since $t \neq t'$ in equation (3.4), her three-dimensional section $t' = \text{const.}$ does *not* coincide with Frank's three-dimensional section $t = \text{const.}$ through their reference frames. The pictures would *look the same*, but they would *not be the same sections through the four-dimensional spacetime*[1].

For later reference we also note that the linearity of the equations (3.5) and (3.6) implies that these transformation laws also hold for coordinate intervals,

$$\Delta x' = \gamma(\Delta x - \beta c\Delta t), \quad c\Delta t' = \gamma(c\Delta t - \beta\Delta x), \tag{3.7}$$

$$\begin{gathered} \Delta y' = \Delta y, \quad \Delta z' = \Delta z, \\ \Delta x = \gamma(\Delta x' + \beta c\Delta t'), \quad c\Delta t = \gamma(ct' + \beta\Delta x'). \end{gathered} \tag{3.8}$$

Substitution into $c^2\Delta t'^2 - \Delta \boldsymbol{x}'^2$ then yields $c^2\Delta t'^2 - \Delta \boldsymbol{x}'^2 = c^2\Delta t^2 - \Delta \boldsymbol{x}^2$, i.e. the Lorentz boost (3.5) and (3.6) does indeed satisfy the requirement (3.1) of leaving the vacuum speed of light unchanged.

3.2 Consequences of Lorentz symmetry

Equations (3.5) and (3.6) are a special case of the general Lorentz transformation introduced in section 3.3, but we can identify any important physical consequences using just the special Lorentz transformations (3.5) and (3.6). We can also neglect the trivially transforming y and z coordinates, i.e. we will focus on events happening at various times t on the x-axis.

To understand the physical consequences of the Lorentz transformations, it is convenient to visualize the transformations in an x, ct-diagram. If we conventionally choose orthogonal x, ct-axis, the x'-axis satisfies $ct' = 0$, and this implies with equation (3.5)

$$ct' = \gamma(ct - \beta x) = 0 \quad \Rightarrow \quad \frac{ct}{x} = \beta, \tag{3.9}$$

[1] If this sounds terribly mysterious at this point (which it must if you have just started to learn relativity), please hang on until we have made it to the spacetime diagram figure 3.2 below.

i.e. the x'-axis (the location of all events happening at time $t' = 0$) is a line with slope β in the x, ct-diagram. On the other hand, the ct'-axis satisfies $x' = 0$:

$$x' = \gamma(x - \beta ct) = 0 \quad \Rightarrow \quad \frac{ct}{x} = \frac{1}{\beta}. \tag{3.10}$$

The ct'-axis is therefore a line with slope $1/\beta$ in the x, ct-diagram. This is shown for $\beta = 0.6$ in figure 3.2.

A diagram depicting both time and location for events is denoted as a *spacetime diagram*, and a continuous line of events through a spacetime diagram is a *world line*. The ct-axis is Frank's world line $x = 0$, while Mary moves along her world line $x' = 0$, i.e. the ct'-axis.

Furthermore, every x' coordinate line $ct' =$ const. would have the same slope β as the x' axis $ct' = 0$, while every ct' coordinate line satisfies $x' =$ const. and therefore has the same slope $1/\beta$ as the ct'-axis $x' = 0$. This leads to the projections of the spacetime event marked with a black dot in figure 3.2 on the corresponding coordinate axes.

We can now also understand the cryptic warning about figure 3.1 as a snapshot of the two reference frames of Frank and Mary. As stated in figure 3.1, it is a snapshot at time t, and therefore lives on a horizontal section $ct =$ const. through figure 3.2. The similar looking picture from Mary's perspective, when Frank is a distance vt' away, lives on an x'-coordinate line $ct' =$ const., i.e. on a line with slope β in figure 3.2. This is a manifestation of the relativity of simultaneity of the two observers.

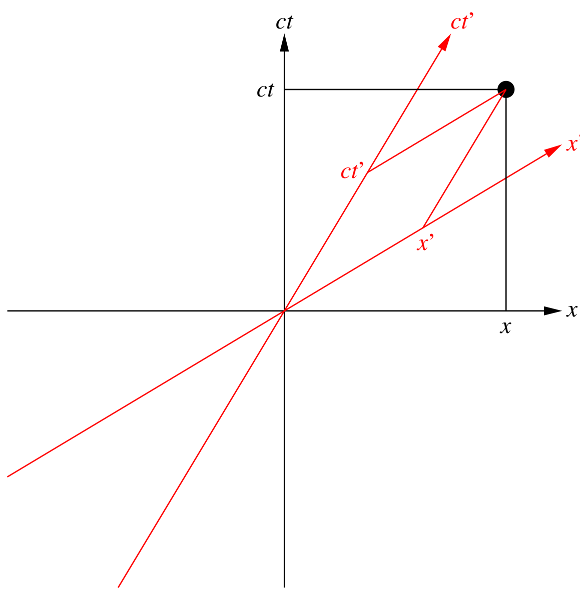

Figure 3.2. The (x, ct) and (x', ct') coordinate systems for $\beta = 0.6$. The (x, ct) axes are conventionally chosen orthogonal. The projections of an event with coordinates (x, ct) and (x', ct') on the corresponding coordinate axes is also shown.

A simultaneous snapshot of the two reference frames from Frank's perspective is not a simultaneous snapshot from Mary's perspective, and vice versa.

We can figure out length and time measurement units in the spacetime diagram figure 3.2 from the observation that the transformations (3.5) leave the coordinate combination $x^2 - c^2t^2$ invariant: $x'^2 - c^2t'^2 = x^2 - c^2t^2$. This gauges the coordinate axes in figure 3.2. The lines $x'^2 - c^2t'^2 = x^2 - c^2t^2 = \text{const.}$ are hyperbolas in the spacetime diagram, and this implies that the point ($x = 0$, $ct = 1$ km) ($\Rightarrow t = 3.33$ μs) and the point ($x' = 0$, $ct' = 1$ km) are located on the intersections of the blue hyperbola $c^2t^2 - x^2 = 1\text{ km}^2$ with the ct-axis and ct'-axis in figure 3.3. The blue hyperbola therefore gauges the time axes in the spacetime diagram in the sense that a clock traveling with Frank along the ct-axis shows $ct = 1$ km at the intersection of the blue hyperbola with the ct-axis, whereas a clock traveling with Mary along the ct'-axis shows $ct' = 1$ km at the intersection of the blue hyperbola with the ct'-axis.

For the gauging of the spatial axes, we note that the points ($x = 1$ km, $ct = 0$) and ($x' = 1$ km, $ct' = 0$) are located on the intersections of the green hyperbolic line $x^2 - c^2t^2 = 1\text{ km}^2$ with the x-axis and x'-axis, respectively.

The angle φ between the x and the x' axis is also the angle between the ct and ct' axis, and is fixed by the relative speed v of the two observers, $\tan\varphi = \beta = v/c$. However, the angles between spacelike and timelike axes (e.g. the angle between the x-axis and the ct-axis) have no physical meaning and are completely arbitrary. E.g. we can also draw the spacetime diagram figure 3.3 with a right angle between the x'-axis and the ct'-axis as in figure 3.4. In that diagram the slope of the x-axis $ct = 0$ is $-\beta$ while the slope of the ct-axis $x = 0$ is $-1/\beta$.

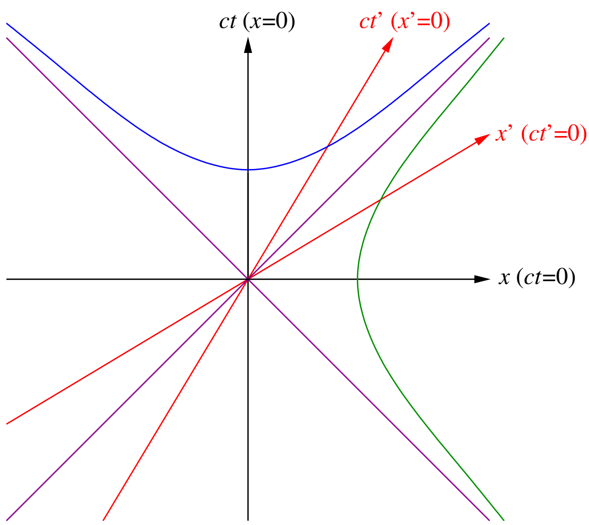

Figure 3.3. The x, ct and x', ct' coordinate systems for $\beta = 0.6$. The blue hyperbolic line satisfies $c^2t'^2 - x'^2 = c^2t^2 - x^2 = 1\text{ km}^2$, the green hyperbolic line satisfies $x'^2 - c^2t'^2 = x^2 - c^2t^2 = 1\text{ km}^2$, and the violet lines are the light cone $c^2t'^2 - x'^2 = c^2t^2 - x^2 = 0$.

For yet another representation we can draw the spacetime diagram without any right angle at all, as e.g. in figure 3.5 where I have chosen a representation which best seems to exhibit the symmetry between the two observers.

The trick for the construction of the symmetric representation in figure 3.5 is to introduce a third coordinate system x'' and ct'', which is not shown in the diagram, but would appear with right angles. Frank moves with speed $c/3$ in negative x'' direction while Mary moves with speed $c/3$ in positive x'' direction in that coordinate

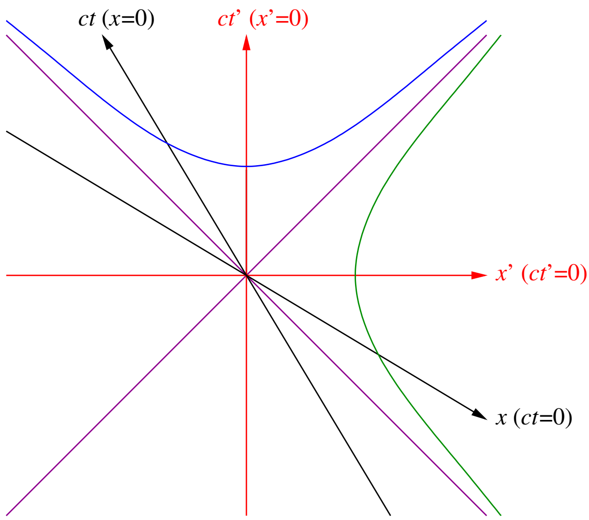

Figure 3.4. The x, ct and x', ct' coordinate systems for $\beta = 0.6$ with a right angle between the x'-axis and the ct'-axis.

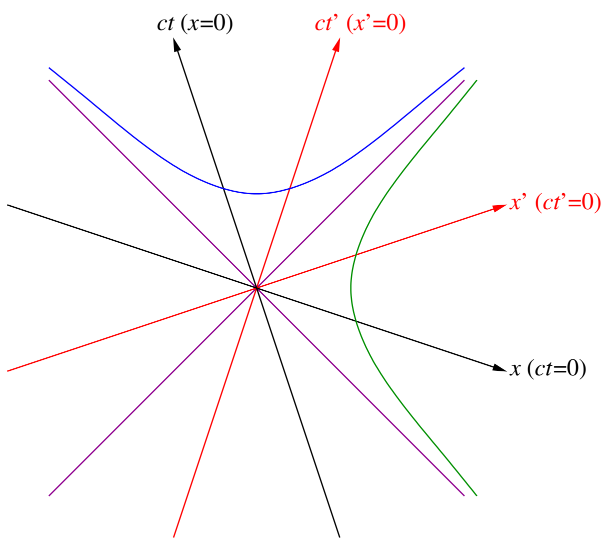

Figure 3.5. The x, ct and x', ct' coordinate systems for $\beta = 0.6$ in a symmetric representation.

system. When we discuss the relativistic addition of velocities in equations (3.16)–(3.18) below, we will find that this implies a relative speed $v = 0.6c$ between Mary and Frank.

Spacetime diagrams are helpful in many situations. They will help us in particular to understand length contraction and time dilation in special relativity.

Length contraction

A stick of length $x = L$ at rest in Frank's frame covers at time $t = 0$ all the points $0 \leqslant x \leqslant L$ on the x-axis in Frank's frame. However, in Mary's frame the stick is a moving stick which extends at time $ct' = 0$ from $x' = 0$ to $x' = L'$. We can calculate L' from the knowledge that the right end of the stick satisfies $x = L$ at all times. This implies that L' is the x' value which corresponds to $x = L$ and $ct' = 0$:

$$x = L,\ ct' = \gamma(ct - \beta L) = 0 \ \Rightarrow\ ct = \beta L \tag{3.11}$$

$$\Rightarrow L' = x'|_{x=L,ct'=0} = x'|_{x=L,ct=\beta L} = L\sqrt{1-\beta^2} < L. \tag{3.12}$$

The moving stick is shorter by a factor $\sqrt{1-\beta^2}$ in Mary's reference frame.

On the other hand, a stick of length L at rest in Mary's frame appears as a moving rod extending at $ct = 0$ from $x = 0$ to

$$L' = x|_{x'=L,ct=0} = x|_{x'=L,ct'=-\beta L} = L\sqrt{1-\beta^2} < L. \tag{3.13}$$

The moving stick is shorter for Frank. This can be visualized in the spacetime diagram in figure 3.6.

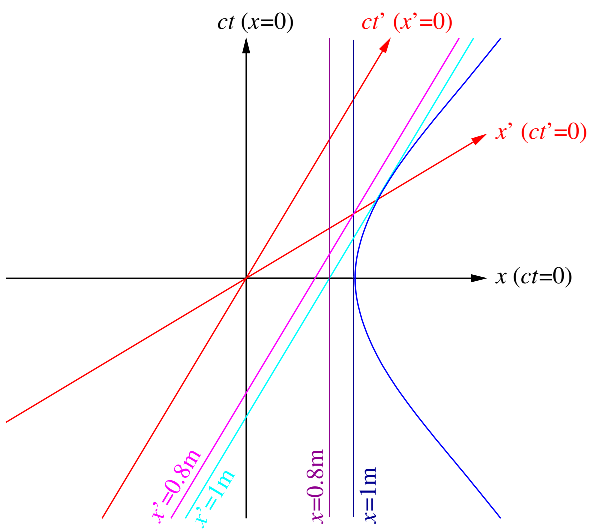

Figure 3.6. The x, ct and x', ct' coordinate systems for $\beta = 0.6$ including the effects of traveling meter sticks and length measurements by Frank and Mary.

The blue hyperbolic line satisfies $x'^2 - c^2t'^2 = x^2 - c^2t^2 = 1\,\mathrm{m}^2$ and gauges the different spatial axes in the spacetime diagram.

The dark blue vertical line is the world line of the right endpoint of Frank's meter stick. It intersects Mary's x'-axis at $x' = 80\,\mathrm{cm}$. Frank's meter stick has a length of 80 cm in Mary's frame. The light blue line is the world line of the right endpoint of Mary's meter stick. It intersects Frank's x-axis at $x = 80\,\mathrm{cm}$. Mary's meter stick has a length of 80 cm in Frank's frame. Note that at each moment of time (where 'moment of time' has different meanings for Frank and Mary) Frank's meter stick consists of different sets of spacetime points for Frank and for Mary. E.g. for $t = 0$ Frank's meter stick consists of all the points between $x = 0$ and $x = 1\,\mathrm{m}$ on the x-axis, while for time $t' = 0$ Frank's meter stick consists of all the points between $x' = 0$ and $x' = 80\,\mathrm{cm}$ on the x'-axis. Similarly, Mary's meter stick at time $t' = 0$ consists of all points between $x' = 0$ and $x' = 1\,\mathrm{m}$ on the x'-axis. At time $t = 0$ Mary's meter stick consists of all the points between $x = 0$ and $x = 80\,\mathrm{cm}$ on the x-axis. The relativity of time implies that Frank and Mary reconstruct the meter sticks as different collections of events in spacetime, and therefore observe different lengths of the meter sticks.

Time dilation

Suppose both Frank and Mary carry copies of identical atomic clocks which tick with a period T. The period $x = 0$, $ct = cT$ on Frank's clock corresponds according to the first equation in (3.5) to a time

$$T' = t'|_{x=0,t=T} = \gamma T = T\Big/\sqrt{1 - \beta^2} > T \tag{3.14}$$

elapsed on Mary's clock. Mary observes that Frank's clock ticks at a slower rate than her own clock. The same applies to Frank's heartbeat and the period of a pendulum traveling with Frank. As described in Mary's reference frame, time in Frank's moving frame is slowed down by a factor γ.

On the other hand, the period $x' = 0$, $ct' = cT$ on Mary's clock corresponds according to the first equation in (3.6) to a time

$$T' = t|_{x'=0,t'=T} = \gamma T > T \tag{3.15}$$

elapsed on Frank's clock. Frank observes that Mary's clock is slowed down by the factor γ. This effect is explained in figure 3.7.

Frank looks at his watch and measures a time $t = 60$ min in the point where the ct-axis (Frank's world line) intersects the blue hyperbola $c^2t^2 - x^2 = c^2t'^2 - x'^2 = c^2 \times 1\,\mathrm{h}^2$. Furthermore, every event on the dark blue horizontal line projects to the time value $t = 60$ min in Frank's coordinate frame, and therefore happens at time $t = 60$ min from his point of view. In particular, in Frank's coordinate system him looking at his watch and reading $t = 60$ min is simultaneous with Mary looking at her watch and reading a time $t' = 48$ min. Mary's watch is slowed down from Frank's point of view. However, in Mary's frame two events are simultaneous if they project onto the same time value t' on her ct'-axis, i.e. if the two events are both on a line with slope β in figure 3.7. In her

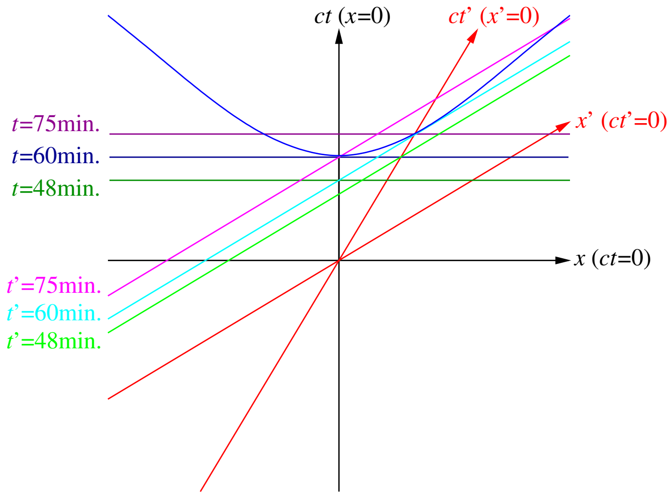

Figure 3.7. The x, ct and x', ct' coordinate systems for $\beta = 0.6$ including the effects of traveling clocks and time measurements by Frank and Mary.

coordinate system Frank reads a time $t = 60$ min from his watch when she reads a time $t' = 75$ min from her watch. Frank's watch is slowed down from Mary's point of view.

The relativity of time implies that Frank looks at his watch and reads 60 minutes, when from his point of view Mary's watch shows a time of 48 min. However, from Mary's point of view Frank looks at his watch and reads 60 minutes when she reads a time of 75 min on her watch.

A clock or an ideal heart under static load will always tick with its designed period Δt in its rest frame, where it does not move, $\Delta \boldsymbol{x} = 0$. However, from the point of view of a moving observer, the clock will move a distance vector $\Delta \boldsymbol{x}'$ in space while it ticks. The fact that the Lorentz transformations (3.7) and (3.8) for coordinate intervals preserve the value of $c^2\Delta t^2 - \Delta \boldsymbol{x}^2$ then implies that relative to the moving observer, the clock ticks with a rate $\Delta t'$ which satisfies $c^2\Delta t^2 = c^2\Delta t'^2 - \Delta \boldsymbol{x}'^2$. This implies $\Delta t' > \Delta t$ for the period of ticking that the moving observer assigns to the clock. A geometric way to interpret this is to say that a clock traveling a distance $|\Delta \boldsymbol{x}'|$ while a time $\Delta t'$ elapses in the frame where $|\Delta \boldsymbol{x}'|$ is measured, measures the *length of its own world line* $c\Delta t = \sqrt{c^2\Delta t'^2 - \Delta \boldsymbol{x}'^2}$.

Relativistic addition of velocities

Suppose Mary observes a spaceship flying at velocity $\boldsymbol{u}' = u_x'\boldsymbol{e}_x + u_y'\boldsymbol{e}_y + u_z'\boldsymbol{e}_z$. Which velocity would Frank assign to this spaceship? We can answer this question by noting that the Lorentz transformation (3.8) from Mary's coordinate intervals to Frank's coordinate intervals implies the velocity components measured by Frank:

$$u_x = \frac{\Delta x}{\Delta t} = \frac{\Delta x' + \beta c \Delta t'}{\Delta t' + \dfrac{\beta}{c}\Delta x'} = \frac{u_x' + v}{1 + (v u_x'/c^2)}, \tag{3.16}$$

$$u_y = \frac{\Delta y}{\Delta t} = \frac{\Delta y'}{\Delta t' + \frac{\beta}{c}\Delta x'}\sqrt{1-\beta^2} = \frac{\sqrt{1-\beta^2}}{1+(vu_x'/c^2)}u_y', \tag{3.17}$$

$$u_z = \frac{\Delta z}{\Delta t} = \frac{\Delta z'}{\Delta t' + \frac{\beta}{c}\Delta x'}\sqrt{1-\beta^2} = \frac{\sqrt{1-\beta^2}}{1+(vu_x'/c^2)}u_z'. \tag{3.18}$$

In particular $u_x' = c/3$ and $v = c/3$ yields $u_x = 0.6c$, which went into the construction of the symmetric spacetime diagram 3.5 for two observers with relative speed $0.6c$.

The transformation laws (3.16)–(3.18) imply also that $\boldsymbol{u}'^2 = c^2$ if and only if $\boldsymbol{u}^2 = c^2$, which again confirms the universality of the vacuum speed of light.

The inverse velocity transformation from Frank's frame to Mary's frame is

$$u_x' = \frac{u_x - v}{1-(vu_x/c^2)}, \tag{3.19}$$

$$u_y' = \frac{\sqrt{1-\beta^2}}{1-(vu_x/c^2)}u_y, \quad u_z' = \frac{\sqrt{1-\beta^2}}{1-(vu_x/c^2)}u_z. \tag{3.20}$$

3.3 The general Lorentz transformation

Equation (3.4) tells us that for motion in a certain direction (x-direction in equation (3.4)), the coordinate in that direction is affected non-trivially by the transformation, while any perpendicular spatial coordinate does not change its value. This allows for a generalization of equation (3.4) to the case that the relative velocity $\boldsymbol{v}$ points in an arbitrary direction.

For convenience we first introduce a rescaled velocity vector $\boldsymbol{\beta} = \boldsymbol{v}/c$. The corresponding unit vector $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}/\beta = \boldsymbol{v}/v$ points in the direction of $\boldsymbol{v}$. It follows from equation (2.34) that the (3×3)-matrix $\hat{\boldsymbol{\beta}} \otimes \hat{\boldsymbol{\beta}}^T$ projects any vector $\boldsymbol{x}$ onto its component parallel to $\boldsymbol{\beta}$

$$\boldsymbol{x}_{\|\beta} = \hat{\boldsymbol{\beta}} \otimes \hat{\boldsymbol{\beta}}^T \cdot \boldsymbol{x}, \tag{3.21}$$

while the component orthogonal to $\boldsymbol{\beta}$ is

$$\boldsymbol{x}_{\perp\beta} = \boldsymbol{x} - \boldsymbol{x}_{\|\beta} = (\underline{\mathbf{1}} - \hat{\boldsymbol{\beta}} \otimes \hat{\boldsymbol{\beta}}^T) \cdot \boldsymbol{x}, \tag{3.22}$$

see also equations (2.37) and (2.38).

The special Lorentz transformation (3.4) tells us that the coordinate $|\boldsymbol{x}_{\|\beta}|$ parallel to $\boldsymbol{v}$ will be rescaled by a factor

$$\gamma = \frac{1}{\sqrt{1 - (v^2/c^2)}} = \frac{1}{\sqrt{1 - \beta^2}}, \tag{3.23}$$

and be shifted by an amount $-\gamma vt = -\gamma\beta ct$. On the other hand, the time coordinate ct will be rescaled by the factor γ and be shifted by an amount $-\gamma\beta|\boldsymbol{x}_{\parallel\beta}|$, while the transverse coordinates $\boldsymbol{x}_{\perp\beta}$ do not change. This leads to the following (4×4)-matrix equation relating the two four-dimensional coordinate vectors:

$$\begin{pmatrix} ct' \\ \boldsymbol{x}' \end{pmatrix} = \begin{pmatrix} \gamma & -\gamma\boldsymbol{\beta}^T \\ -\gamma\boldsymbol{\beta} & \underline{1} - \hat{\boldsymbol{\beta}} \otimes \hat{\boldsymbol{\beta}}^T + \gamma\hat{\boldsymbol{\beta}} \otimes \hat{\boldsymbol{\beta}}^T \end{pmatrix} \cdot \begin{pmatrix} ct \\ \boldsymbol{x} \end{pmatrix} \tag{3.24}$$

This is the general Lorentz transformation between two observers if the origins of their four-dimensional reference frames coincide, i.e. $ct' = x' = y' = z' = 0 \Leftrightarrow ct = x = y = z = 0$, if the spatial axes of their coordinate frames *appear* to coincide in the (ct, x, y, z) system at $t = 0$ and in the (ct', x', y', z') system at $t' = 0$, and if the time direction is not reversed. We can generalize this by allowing for constant shifts of the coordinates and for a rotation of the spatial axes,

$$\begin{pmatrix} ct' \\ \boldsymbol{x}' \end{pmatrix} = \begin{pmatrix} \gamma & -\gamma\boldsymbol{\beta}^T \\ -\gamma\boldsymbol{\beta} & \underline{1} - \hat{\boldsymbol{\beta}} \otimes \hat{\boldsymbol{\beta}}^T + \gamma\hat{\boldsymbol{\beta}} \otimes \hat{\boldsymbol{\beta}}^T \end{pmatrix} \cdot \begin{pmatrix} 1 & \boldsymbol{0}^T \\ \boldsymbol{0} & \underline{R} \end{pmatrix} \cdot \begin{pmatrix} ct - cT \\ \boldsymbol{x} - \boldsymbol{X} \end{pmatrix}. \tag{3.25}$$

Here $\underline{R}$ is a 3×3 rotation matrix which can also include reflections of spatial axes. Without the coordinate shifts this is the most general orthochronous (i.e. time-direction preserving) Lorentz transformation. With the coordinate shifts included it is denoted as an *inhomogeneous Lorentz transformation* or a *Poincaré transformation*.

We can express equation (3.25) succinctly in the form

$$x'^{\mu} = \Lambda^{\mu}{}_{\nu}(x^{\nu} - X^{\nu}). \tag{3.26}$$

where $x^{\mu} = (ct, \boldsymbol{x})$ is the *4-vector* of coordinates and the (4×4) transformation matrix Λ is

$$\Lambda = \{\Lambda^{\mu}{}_{\nu}\} = \begin{pmatrix} \gamma & -\gamma\boldsymbol{\beta}^T \\ -\gamma\boldsymbol{\beta} & \underline{1} - \hat{\boldsymbol{\beta}} \otimes \hat{\boldsymbol{\beta}}^T + \gamma\hat{\boldsymbol{\beta}} \otimes \hat{\boldsymbol{\beta}}^T \end{pmatrix} \cdot \begin{pmatrix} 1 & \boldsymbol{0}^T \\ \boldsymbol{0} & \underline{R} \end{pmatrix}. \tag{3.27}$$

Up to a possible time reflection $t \to -t$ equation (3.25) is the general coordinate transformation $\{ct, \boldsymbol{x}\} = \{ct, x, y, z\} \to \{ct', \boldsymbol{x}'\} = \{ct', x', y', z'\}$ which leaves the expression $\Delta\boldsymbol{x}^2 - c^2\Delta t^2$ invariant, i.e. such that for arbitrary coordinate differentials $c\Delta t$, $\Delta\boldsymbol{x}$ we have

$$\Delta\boldsymbol{x}^2 - c^2\Delta t^2 = \Delta\boldsymbol{x}'^2 - c^2\Delta t'^2. \tag{3.28}$$

This equation implies in particular that an observer in the (ct, x, y, z) reference frame sees a light wave moving at speed c: $\Delta\boldsymbol{x}^2 - c^2\Delta t^2 = 0$, if and only if this light wave also moves with speed c for an observer in the (ct', x', y', z') reference frame: $\Delta\boldsymbol{x}'^2 - c^2\Delta t'^2 = 0$.

More details on 4-vectors

Constant offsets X^μ between coordinate systems vanish for differences Δx^μ of coordinates. Equation (3.26) therefore implies $\Delta x^\mu = \Lambda^\mu{}_\alpha \Delta x^\alpha$ for a Lorentz transformation of coordinate differences. This yields for differentials the transformation law

$$dx'^\mu = \Lambda^\mu{}_\alpha dx^\alpha. \tag{3.29}$$

The condition (3.28) becomes

$$d\mathbf{x}^2 - c^2dt^2 = d\mathbf{x}'^2 - c^2dt'^2, \tag{3.30}$$

and this can also be written as

$$\eta_{\mu\nu}dx'^\mu dx'^\nu = \eta_{\alpha\beta}dx^\alpha dx^\beta. \tag{3.31}$$

This defines the special (4×4)-matrix (including the 3×3 unit matrix $\underline{1}$ as a submatrix)

$$\{\eta_{\mu\nu}\} = \begin{pmatrix} -1 & \mathbf{0}^T \\ \mathbf{0} & \underline{1} \end{pmatrix}, \tag{3.32}$$

which is denoted as the *Minkowski metric*. The equation (2.7) for a metric becomes

$$\eta_{\mu\alpha}\eta^{\alpha\nu} = \eta_\mu{}^\nu = \delta_\mu{}^\nu, \tag{3.33}$$

where the inverse Minkowski metric $(\eta^{-1})^{\alpha\nu} = \eta^{\alpha\nu}$ (see equation (2.6)) has exactly the same matrix elements as the Minkowski metric.

The substitution $dx \to dx + dy$ in equation (3.31) shows that equation (3.28) also implies

$$\eta_{\mu\nu}dx'^\mu dy'^\nu = \eta_{\alpha\beta}dx^\alpha dy^\beta \tag{3.34}$$

for any pair of Lorentz transformed 4-vectors dx and dy. Lorentz transformations must therefore leave the Minkowski metric $\eta_{\mu\nu}$ invariant:

$$\eta_{\mu\nu}\Lambda^\mu{}_\alpha\Lambda^\nu{}_\beta = \eta_{\alpha\beta}, \tag{3.35}$$

and if we multiply this equation with the components $\eta^{\beta\gamma}$ of the inverse Minkowski metric, we find

$$\eta_{\mu\nu}\Lambda^\mu{}_\alpha\Lambda^\nu{}_\beta\eta^{\beta\gamma} = \delta_\alpha{}^\gamma \equiv \eta_\alpha{}^\gamma. \tag{3.36}$$

This provides us with a relation between the (4×4)-matrix Λ with 'pulled indices' and its inverse matrix Λ^{-1}:

$$\Lambda_\mu{}^\gamma \equiv \eta_{\mu\nu}\Lambda^\nu{}_\beta\eta^{\beta\gamma} = (\Lambda^{-1})^\gamma{}_\mu. \tag{3.37}$$

Explicitly, if the Lorentz transformation matrix with 'canonical' index position is

$$\{\Lambda^{\mu}{}_{\nu}\} = \begin{pmatrix} \gamma & -\gamma\boldsymbol{\beta}^T \\ -\gamma\boldsymbol{\beta} & \underline{1} - \hat{\boldsymbol{\beta}} \otimes \hat{\boldsymbol{\beta}}^T + \gamma\hat{\boldsymbol{\beta}} \otimes \hat{\boldsymbol{\beta}}^T \end{pmatrix} \cdot \begin{pmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & \underline{R} \end{pmatrix} \tag{3.38}$$

then the matrix with pulled indices is

$$\{\Lambda_{\mu}{}^{\nu}\} \equiv \{\eta_{\mu\rho}\Lambda^{\rho}{}_{\sigma}\eta^{\sigma\nu}\} = \begin{pmatrix} \gamma & \gamma\boldsymbol{\beta}^T \\ \gamma\boldsymbol{\beta} & \underline{1} - \hat{\boldsymbol{\beta}} \otimes \hat{\boldsymbol{\beta}}^T + \gamma\hat{\boldsymbol{\beta}} \otimes \hat{\boldsymbol{\beta}}^T \end{pmatrix} \cdot \begin{pmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & \underline{R} \end{pmatrix} \tag{3.39}$$

and the inverse matrix is

$$\begin{aligned} \{(\Lambda^{-1})^{\nu}{}_{\mu}\} &= \begin{pmatrix} 1 & \mathbf{0}^T \\ \mathbf{0} & \underline{R}^T \end{pmatrix} \cdot \begin{pmatrix} \gamma & \gamma\boldsymbol{\beta}^T \\ \gamma\boldsymbol{\beta} & \underline{1} - \hat{\boldsymbol{\beta}} \otimes \hat{\boldsymbol{\beta}}^T + \gamma\hat{\boldsymbol{\beta}} \otimes \hat{\boldsymbol{\beta}}^T \end{pmatrix} \\ &= \{\Lambda_{\mu}{}^{\nu}\}^T \equiv \{(\Lambda^T)^{\nu}{}_{\mu}\}. \end{aligned} \tag{3.40}$$

Equations (3.38)–(3.40) show that index positions matter as soon as a boost is involved, $\boldsymbol{\beta} \neq 0$, because

$$\Lambda^{\mu}{}_{\nu}(\boldsymbol{\beta}) = \Lambda_{\nu}{}^{\mu}(-\boldsymbol{\beta}) \neq \Lambda_{\nu}{}^{\mu}(\boldsymbol{\beta}). \tag{3.41}$$

If a matrix Λ is given without specifying the index positions, the default convention is that the matrix (3.38) is meant.

We can also 'pull' or 'draw' indices on 4-vectors, e.g. $dx_{\alpha} \equiv \eta_{\alpha\beta}dx^{\beta} = (-cdt, d\mathbf{x})$. Equation (3.37) implies the following transformation law for this 4-vector under the Lorentz transformation (3.29):

$$\begin{aligned} dx'_{\mu} = \eta_{\mu\nu}dx'^{\nu} &= \eta_{\mu\nu}\Lambda^{\nu}{}_{\alpha}dx^{\alpha} = \eta_{\mu\nu}\Lambda^{\nu}{}_{\alpha}\eta^{\alpha\beta}dx_{\beta} \\ &= \Lambda_{\mu}{}^{\beta}dx_{\beta} = dx_{\beta}(\Lambda^{-1})^{\beta}{}_{\mu}. \end{aligned} \tag{3.42}$$

4-vectors with this kind of transformation behavior are denoted as *covariant 4-vectors*, while dx^{μ} is an example of a *contravariant 4-vector*. This agrees with the corresponding definitions in chapter 2, with the transformation matrix Λ corresponding to the matrix $\underline{M}^{-1}$ in chapter 2.

Another example of a covariant 4-vector is the vector of partial derivatives in four dimensions,

$$\partial_{\mu} \equiv \frac{\partial}{\partial x^{\mu}} = \left(\frac{1}{c}\frac{\partial}{\partial t}, \nabla\right). \tag{3.43}$$

We can prove that this is indeed a covariant 4-vector using the chain rule for differentiation,

$$\partial'_{\mu} \equiv \frac{\partial}{\partial x'^{\mu}} = \frac{\partial x^{\alpha}}{\partial x'^{\mu}}\frac{\partial}{\partial x^{\alpha}}. \tag{3.44}$$

However, we have $dx^{\alpha} = (\Lambda^{-1})^{\alpha}{}_{\nu}dx'^{\nu}$ and therefore $\partial x^{\alpha}/\partial x'^{\mu} = (\Lambda^{-1})^{\alpha}{}_{\mu}$, and this implies

$$\partial'_\mu = (\Lambda^{-1})^\alpha{}_\mu \partial_\alpha = \Lambda_\mu{}^\alpha \partial_\alpha, \tag{3.45}$$

as expected for a covariant 4-vector.

Since co- and contravariant indices transform with mutually inverse transformation matrices, pairs of those indices do not transform if they are summed over. E.g. the term $\partial_\alpha F^{\alpha\beta}$ transforms under Lorentz transformations according to

$$\partial'_\mu F'^{\mu\nu} = \Lambda_\mu{}^\alpha \partial_\alpha \Lambda^\mu{}_\beta \Lambda^\nu{}_\gamma F^{\beta\gamma} = \Lambda^\nu{}_\gamma \eta^\alpha{}_\beta \partial_\alpha F^{\beta\gamma} = \Lambda^\nu{}_\gamma \partial_\alpha F^{\alpha\gamma}, \tag{3.46}$$

i.e. the summed index pair does not contribute to the transformation. Only 'free' (= unsummed) indices transform under the Lorentz transformation.

Chapter 4

Relativistic dynamics

We have already heard that the basic symmetries of electromagnetism and Newtonian mechanics are mutually incompatible. Newton's equations of motion hold in different reference frames if those frames are related by Galilei transformations, whereas the wave equations of electromagnetism are only preserved between different observers if their coordinates are related through Lorentz transformations. We have seen the Galilei invariance of Newton's equation in section 1.1, and we will see the Lorentz invariance of the wave equations of electrodynamics in section 4.2. So who was wrong (or rather, not entirely correct): Newton or Maxwell? The Michelson experiment convinced us that different inertial observers must be related by Lorentz transformations, i.e. Newton's equation can only be an approximation for small particle speeds $v \ll c$ and needs to be modified to comply with Lorentz transformations. We therefore need to embark on a quest to find the relativistic generalization of Newton's equation.

4.1 Energy–momentum vectors and the relativistic Newton equation

In STR it is advantageous to express everything in quantities which transform linearly with combinations of the Lorentz transformation matrices Λ (3.27). For example, conservation laws must hold not only in one inertial frame, but in all inertial frames, and if we can find a definition of energy and momentum which transforms linearly under Lorentz transformations, then the linearity of the transformation ensures the preservation of energy and momentum conservation in transitions between different inertial frames. We will see this explicitly in equations (4.18) and (4.19).

A first step towards the identification of linearly transforming energy and momentum vectors involves the concepts of *eigentime* and 4-velocity.

doi:10.1088/2053-2571/aaf173ch4

Proper time ('eigentime') and 4-velocity

The transformation law

$$dx'^{\mu} = \Lambda^{\mu}{}_{\nu} dx^{\nu}, \tag{4.1}$$

implies that ordinary velocities $d\mathbf{x}/dt$ and accelerations $d^2\mathbf{x}/dt^2$ do not transform linearly under Lorentz boosts because the time coordinate in the denominators is also transformed. We have found the nonlinear transformation behavior of the velocity components explicitly in equations (3.16)–(3.20). This nonlinear transformation behavior of the ordinary velocity components can be avoided by substituting the physical velocities and accelerations with 'proper' velocities and accelerations, which do not require division by a transforming time parameter t. The trick is to use the *proper time* or *eigentime* of a moving object in the denominator instead of the lab time t.

To explain the notion of *eigentime*, we consider a moving object which is moving with momentary velocity $\boldsymbol{v} = d\mathbf{x}/dt$ relative to our (ct, x) frame, but is at rest in its own $(ct', \mathbf{x}')$ reference frame, i.e. in its own frame the trajectory of the object is $\mathbf{x}' = 0$. We know that the Lorentz transformation (4.1) leaves the product $dx^{\mu}dx_{\mu}$ invariant,

$$dx'^{\mu}dx'_{\mu} = d\mathbf{x}'^2 - c^2dt'^2 = dx^{\mu}dx_{\mu} = d\mathbf{x}^2 - c^2dt^2. \tag{4.2}$$

This implies that the time $dt' \equiv d\tau$ measured by the moving object along its own path $\mathbf{x}' = 0$ satisfies

$$d\tau^2 = dt^2 - \frac{1}{c^2}d\mathbf{x}^2 = \left(1 - \frac{v^2}{c^2}\right)dt^2, \tag{4.3}$$

and therefore we have up to an integration constant

$$\tau = \int dt\sqrt{1 - (v^2/c^2)} = \int \frac{dt}{\gamma}. \tag{4.4}$$

This is an *invariant*, i.e. it has *the same value for each observer*. Every observer uses their own clocks and measures their own specific time interval Δt between any two events happening to the moving object, but all observers agree on the same value

$$\Delta\tau = \int_0^{\Delta t} dt\sqrt{1 - (v^2/c^2)} \tag{4.5}$$

which elapsed on a clock moving with the object. This time $\Delta\tau$ measured by an object between any two events happening to the object is the *proper time* or *eigentime* of the object.

The definition of proper time motivates a corresponding definition of the *proper velocity* or *eigenvelocity* of an object in an observer's reference frame: We define the proper velocity by dividing the change in the object's coordinates $d\mathbf{x}$ *in the observer's reference frame* by the time interval $d\tau$ elapsed *in the object's reference frame* while it was moving by $d\mathbf{x}$:

$$\boldsymbol{U} = \frac{d\boldsymbol{x}}{d\tau} = \gamma\frac{d\boldsymbol{x}}{dt} = \gamma\boldsymbol{v} = \frac{\boldsymbol{v}}{\sqrt{1-(v^2/c^2)}}. \tag{4.6}$$

This is a hybrid construction in the sense that a set of coordinate intervals $d\boldsymbol{x}$ measured in the observer's frame is divided by a coordinate interval $d\tau$ measured in the object's frame[1].

The notion of proper velocity is useful because it can be extended to a 4-vector due to the fact that $\{dx^\mu\} = (dx^0, d\boldsymbol{x}) = (cdt, d\boldsymbol{x})$ is a 4-vector under Lorentz transformations. We can use this to define a notion of velocity which is also a 4-vector. We simply combine the definition (4.6) for the spatial components of the eigenvelocity with the corresponding timelike component,

$$U^0 = \frac{dx^0}{d\tau} = \frac{cdt}{d\tau} = \gamma c. \tag{4.7}$$

This leads to the full relativistic eigenvelocity or *4-velocity* of the moving object in the form

$$(U^0, \boldsymbol{U}) = \gamma(c, \boldsymbol{v}). \tag{4.8}$$

The 4-velocity satisfies

$$U^2 \equiv U^\mu U_\mu \equiv \eta_{\mu\nu}U^\mu U^\nu = \boldsymbol{U}^2 - (U^0)^2 = -c^2. \tag{4.9}$$

For future reference we also note that solving equation (4.6) for $\boldsymbol{v}$ yields

$$\boldsymbol{v} = \frac{\boldsymbol{U}}{\gamma} = \frac{\boldsymbol{U}}{\sqrt{1+(\boldsymbol{U}^2/c^2)}} = c\frac{\boldsymbol{U}}{U^0}, \tag{4.10}$$

and the γ factor in terms of 4-velocity components is

$$\gamma = U^0\big/c = \sqrt{1+(\boldsymbol{U}^2/c^2)}. \tag{4.11}$$

Relativistic momentum and energy

The nonlinear transformation behavior (3.16)–(3.20) of the ordinary velocity components between different lab frames implies that the conservation laws

$$\sum_i \boldsymbol{p}_i^{(\mathrm{in})} = \sum_i \boldsymbol{p}_i^{(\mathrm{out})}, \quad \sum_i E_i^{(\mathrm{in})} = \sum_i E_i^{(\mathrm{out})} \tag{4.12}$$

for total momentum and energy in a collision of particles would not be preserved under Lorentz transformations if we use the non-relativistic definitions for momentum and energy. This would imply that if momentum and energy conservation

[1] We will see below that there is a limit $v \leqslant c$ on the *physical speed* $v = |\boldsymbol{v}|$ of moving objects. No such limit holds for the 'eigenspeed' $|\boldsymbol{U}|$. However, the speed of signal transmission relative to an observer is v, not $|\boldsymbol{U}|$, i.e. the concept of eigenvelocity does not invalidate the absolute speed limit on signal transmission in any practical way.

would hold for one observer, they would not hold for another observer in a different lab frame with different velocity!

However, the conservation laws would hold in all reference frames if energy and momentum transform *linearly*, i.e. like a 4-vector, under Lorentz transformations. We have already identified velocity components $\{U^\mu\} = \gamma(c, \boldsymbol{v})$ which transform like a 4-vector and also have the property $\lim_{\beta\to 0}\boldsymbol{U} = \boldsymbol{v}$. This observation motivates the definition of the *4-momentum*

$$p^0 = mU^0, \quad \boldsymbol{p} = m\boldsymbol{U}, \tag{4.13}$$

i.e. the relativistic definition of the spatial momentum of a particle of mass m and physical velocity $\boldsymbol{v}$ is

$$\boldsymbol{p} = m\boldsymbol{U} = \gamma m\boldsymbol{v} = \frac{m\boldsymbol{v}}{\sqrt{1-(v^2/c^2)}}. \tag{4.14}$$

The timelike momentum component

$$p^0 = mU^0 = \gamma mc = \frac{mc}{\sqrt{1-(v^2/c^2)}} \tag{4.15}$$

must apparently be related to the fourth conserved quantity in collisions, i.e. to energy. We can confirm this by studying the non-relativistic limit: $v \ll c$ yields

$$p^0 \approx mc\left(1 + \frac{v^2}{2c^2}\right) = mc + \frac{1}{c}K, \tag{4.16}$$

where $K = mv^2/2$ is the non-relativistic kinetic energy of the particle. The quantity cp^0 is therefore the energy of a particle of mass m and speed v:

$$E = cp^0 = \gamma mc^2 = \frac{mc^2}{\sqrt{1-(v^2/c^2)}}. \tag{4.17}$$

The linear transformation under the Lorentz transformation (4.1), $p'^\mu = \Lambda^\mu{}_\nu p^\nu$, then implies preservation of energy and momentum conservation under Lorentz transformations: Suppose that $p^\mu_{(\text{in})}$ and $p^\mu_{(\text{out})}$ are the total energy–momentum 4-vectors of a particle system before and after a scattering event, respectively. The transformation law

$$p'^\mu_{(\text{out})} - p'^\mu_{(\text{in})} = \Lambda^\mu{}_\nu\left(p^\nu_{(\text{out})} - p^\nu_{(\text{in})}\right) \tag{4.18}$$

implies covariance of energy–momentum conservation,

$$p^\mu_{(\text{out})} = p^\mu_{(\text{in})} \Leftrightarrow p'^\mu_{(\text{out})} = p'^\mu_{(\text{in})}. \tag{4.19}$$

Energy–momentum conservation does not depend on the reference frame.

Note that the identification of energy and momentum as components of the 4-momentum vector $p^\mu = \{E/c, \boldsymbol{p}\}$ also implies that the distinction between energy and spatial momentum is as observer dependent as the distinction between time and space, e.g. under a Lorentz boost (3.24) with velocity $c\boldsymbol{\beta} = \boldsymbol{v}_O = v_O\boldsymbol{e}_x$, $\gamma = [1 - (v_O/c)^2]^{-1/2}$,

$$E' = \gamma(E - v_O p_x), \quad p_x' = \gamma\Big[p_x - \Big(v_O E/c^2\Big)\Big], \quad p_y' = p_y, \quad p_z' = p_z. \tag{4.20}$$

Note that $\boldsymbol{v}_O$ is the relative velocity between the two inertial frames which measure energy and momentum of the particle, i.e. $\boldsymbol{v}_O$ is generically different from the particle velocity $\boldsymbol{v}$ relative to the first reference frame or the particle velocity $\boldsymbol{v}'$ relative to the second reference frame.

Relativistic dispersion relations

We can get an expression for the velocity from the division of the equations (4.14) and (4.17),

$$\boldsymbol{v} = c^2\boldsymbol{p}/E. \tag{4.21}$$

On the other hand, a relation between energy and momentum is also called a dispersion relation, and subtracting squares yields the relativistic dispersion relation

$$E^2 - c^2\boldsymbol{p}^2 = m^2c^4 \quad \Rightarrow \quad E^2 = m^2c^4 + c^2\boldsymbol{p}^2. \tag{4.22}$$

This dispersion relation is usually written in the form

$$p^2 + m^2c^2 = 0, \tag{4.23}$$

where $p^2 \equiv p^\mu p_\mu \equiv \eta_{\mu\nu}p^\mu p^\nu = \boldsymbol{p}^2 - (E/c)^2$ is the square of the momentum 4-vector $\{p^\mu\} = (E/c, \boldsymbol{p})$.

Equation (4.22) implies in particular for $m = 0$ the relation $E = cp$, and substitution into equation (4.21) yields $v = c$, i.e. massless particles in vacuum move at speed c.

Sometimes it is also useful to express velocity or speed only as functions of momentum or energy,

$$\boldsymbol{v} = \frac{c\boldsymbol{p}}{\sqrt{m^2c^2 + p^2}}, \quad v = \frac{cp}{\sqrt{m^2c^2 + p^2}} = \frac{c}{E}\sqrt{E^2 - m^2c^4}. \tag{4.24}$$

These equations tell us that a massive particle always has $v < c$, or equivalently that it would require an infinite amount of energy to accelerate a massive particle to the vacuum speed of light. For massless particles we recover that $v = c$ in vacuum, irrespective of the momentum p and energy E of the particle.

The special relativistic version of Newton's equation

The rate of change of 4-momentum with eigentime defines a force 4-vector or *4-force*

$$f^\mu = \frac{d}{d\tau}m\frac{dx^\mu}{d\tau} = \frac{dp^\mu}{d\tau} = \gamma\frac{d}{dt}p^\mu. \tag{4.25}$$

This behaves like a 4-vector, i.e. transforms like $f'^{\mu} = \Lambda^{\mu}{}_{\nu} f^{\nu}$, because we divided the 4-vector dp^{μ} by the invariant eigentime interval $d\tau$.

A three-dimensional force of the form

$$\boldsymbol{F} = \frac{d}{dt}\boldsymbol{p} = \frac{1}{\gamma}\boldsymbol{f} \tag{4.26}$$

is still commonly used, but $\boldsymbol{F}$ is *not* the spatial component of a 4-vector, whereas $\boldsymbol{f} = \gamma \boldsymbol{F}$ is.

The relativistic dispersion relation $E = c\sqrt{\boldsymbol{p}^2 + m^2c^2}$ implies for the 0-component f^0

$$\begin{aligned} f^0 = \frac{d}{d\tau}m\frac{dx^0}{d\tau} &= \frac{d}{d\tau}\left(m\gamma\frac{dx^0}{dt}\right) = \frac{d}{d\tau}(\gamma mc) = \frac{d}{d\tau}\frac{E}{c} \\ &= \frac{d}{d\tau}\sqrt{\boldsymbol{p}^2 + m^2c^2} \\ &= \frac{\boldsymbol{p}}{\sqrt{\boldsymbol{p}^2 + m^2c^2}} \cdot \frac{d}{d\tau}\boldsymbol{p} = \frac{\boldsymbol{v}}{c} \cdot \frac{d}{d\tau}\boldsymbol{p} = \frac{\boldsymbol{v}}{c} \cdot \boldsymbol{f}. \end{aligned} \tag{4.27}$$

The force 4-vector f^{μ} can therefore be expressed in terms of $\boldsymbol{F}$ in the form

$$(f^0, \boldsymbol{f}) = (\boldsymbol{\beta} \cdot \boldsymbol{f}, \boldsymbol{f}) = (\gamma\boldsymbol{\beta} \cdot \boldsymbol{F}, \gamma\boldsymbol{F}). \tag{4.28}$$

Multiplication of the equation $dp^0/d\tau = f^0$ in the form (cf equation (4.27))

$$\frac{d}{d\tau}\frac{E}{c} = \frac{\boldsymbol{v}}{c} \cdot \boldsymbol{f} \tag{4.29}$$

with c/γ yields the conventional expression for energy balance,

$$\frac{d}{dt}E = \frac{c}{\sqrt{\boldsymbol{p}^2 + m^2c^2}}\boldsymbol{p} \cdot \frac{d}{dt}\boldsymbol{p} = \boldsymbol{v} \cdot \frac{d}{dt}\boldsymbol{p} = \boldsymbol{v} \cdot \boldsymbol{F}. \tag{4.30}$$

In summary, we conclude that the special relativistic version of Newton's equation is $dp^{\mu}/d\tau = f^{\mu}$, where f^{μ} is related to the lab frame force $\boldsymbol{F}$ through equation (4.28).

4.2 The manifestly covariant formulation of electrodynamics

Lorentz or Poincaré transformations (3.25) were essentially discovered for the first time by Voigt in 1887 in his analysis of the electromagnetic wave equations[2]. We can easily understand the relativistic invariance of wave equations from what we have already learned about 4-vectors. The relativistic wave operator is

$$\Delta - \frac{1}{c^2}\frac{\partial^2}{\partial t^2} = \partial_{\mu}\partial^{\mu}, \tag{4.31}$$

[2] See Voigt W 1887 *Nachrichten von der Königlichen Gesellschaft der Wissenschaften und der Georg-Augusts-Universität zu Göttingen* 41.

and since summations over co- and contravariant indices are Lorentz invariant, the wave operator has the same form in every inertial frame, $\partial'_\mu \partial'^\mu = \partial_\mu \partial^\mu$. However, what about the full set of Maxwell's equations, from which the wave equations for electromagnetic waves are derived? When we look at Maxwell's equations in standard notation,

$$\nabla \cdot \boldsymbol{E} = \frac{1}{\epsilon_0}\varrho, \quad \nabla \times \boldsymbol{B} - \frac{1}{c^2}\frac{\partial}{\partial t}\boldsymbol{E} = \mu_0 \boldsymbol{j} \tag{4.32}$$

$$\nabla \cdot \boldsymbol{B} = 0, \quad \nabla \times \boldsymbol{E} + \frac{\partial}{\partial t}\boldsymbol{B} = \boldsymbol{0}, \tag{4.33}$$

Lorentz invariance is far from obvious. Besides the 4-vector of spacetime coordinates $x^\mu = (ct, \boldsymbol{x})$, there appear three three-dimensional vectors $\boldsymbol{E}$, $\boldsymbol{B}$ and $\boldsymbol{j}$, and a scalar ϱ. Knowing that $\boldsymbol{j}$ is a vector under spatial rotations while ϱ is a scalar, and taking into account that both quantities describe the motion of electric charges through spacetime, easily enough leads to the identification of a current 4-vector $j^\mu = (\varrho c, \boldsymbol{j})$. This idea is confirmed by the observation that with this identification, the conservation for electric charge can be written in manifestly Lorentz invariant form[3],

$$\frac{\partial \varrho}{\partial t} + \nabla \cdot \boldsymbol{j} = \partial_\mu j^\mu = 0. \tag{4.34}$$

The expression $\partial_\mu j^\mu$ is manifestly Lorentz invariant, because we have seen that Lorentz transformations cancel between contracted upper and lower indices: The transformation equations

$$x'^\mu = \Lambda^\mu{}_\nu x^\nu \quad \Rightarrow \quad \partial'_\mu = \Lambda_\mu{}^\nu \partial_\nu, \quad j'^\mu(x') = \Lambda^\mu{}_\nu j^\nu(x), \tag{4.35}$$

yield $\partial'_\mu j'^\mu(x') = \partial_\mu j^\mu(x)$. Therefore charge is conserved in every reference frame if it is conserved in one reference frame, and in every inertial reference frame charge conservation is expressed through the equation (4.34).

However, what about the three-dimensional vectors $\boldsymbol{E}$ and $\boldsymbol{B}$? There are certainly no remaining scalars anywhere in Maxwell's equations (4.32) and (4.33) that we could combine with $\boldsymbol{E}$ and $\boldsymbol{B}$ to promote those vectors to 4-vectors. The problem of the correct relativistic identification of the electromagnetic field strengths was solved by Hermann Minkowski in 1907[4]. Recall that the homogeneous Maxwell's equations are solved through potentials Φ and $\boldsymbol{A}$:

$$\boldsymbol{B} = \nabla \times \boldsymbol{A}, \quad \boldsymbol{E} = -\frac{\partial}{\partial t}\boldsymbol{A} - \nabla\Phi. \tag{4.36}$$

Combining the potentials into a 4-vector of potentials (or *4-potential*) $A_\mu = \{-\Phi/c, \boldsymbol{A}\}$, shows that the electromagnetic field strengths E_i, B_i are related to antisymmetric combinations of the 4-vectors ∂_μ, A_ν,

[3] If you are not familiar from electrodynamics with the local formulation (4.34) of charge conservation, you can briefly jump ahead to the discussion leading from equation (4.85) to equation (4.88).

[4] Minkowski H 1910 *Mathematische Annalen* 68 472. A translation of his results for dielectric materials into contemporary tensor notation can be found in Dick R 2009 *Ann. Phys.* 18 174.

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu = \begin{pmatrix} 0 & -E_1/c & -E_2/c & -E_3/c \\ E_1/c & 0 & B_3 & -B_2 \\ E_2/c & -B_3 & 0 & B_1 \\ E_3/c & B_2 & -B_1 & 0 \end{pmatrix}, \tag{4.37}$$

i.e. they are the six independent components of an antisymmetric 4×4 *electromagnetic field strength tensor F*. The matrix elements $F_{\mu\nu}$ give the tensor in covariant components. Pulling indices with the Minkowski metric yields the contravariant components of F,

$$F^{\mu\nu} = \eta^{\mu\alpha}\eta^{\nu\beta}F_{\alpha\beta} = \partial^\mu A^\nu - \partial^\nu A^\mu = \begin{pmatrix} 0 & E_1/c & E_2/c & E_3/c \\ -E_1/c & 0 & B_3 & -B_2 \\ -E_2/c & -B_3 & 0 & B_1 \\ -E_3/c & B_2 & -B_1 & 0 \end{pmatrix}. \tag{4.38}$$

We can now easily determine the transformation behavior of the electromagnetic fields under Lorentz transformations, $x^\mu \to x'^\mu = \Lambda^\mu{}_\alpha x^\alpha$. The equations

$$\partial_\mu \to \partial'_\mu = \Lambda_\mu{}^\alpha \partial_\alpha, \quad A_\mu(x) \to A'_\mu(x') = \Lambda_\mu{}^\alpha A_\alpha(x), \tag{4.39}$$

imply

$$\begin{aligned} F_{\mu\nu}(x) \to F'_{\mu\nu}(x') &= \partial'_\mu A'_\nu(x') - \partial'_\nu A'_\mu(x') = \Lambda_\mu{}^\alpha \Lambda_\nu{}^\beta[\partial_\alpha A_\beta(x) - \partial_\beta A_\alpha(x)] \\ &= \Lambda_\mu{}^\alpha \Lambda_\nu^\beta F_{\alpha\beta}(x). \end{aligned} \tag{4.40}$$

Evaluation of equation (4.40) for a boost

$$\{\Lambda^\mu{}_\nu\} = \begin{pmatrix} \gamma & -\gamma\boldsymbol{\beta}^T \\ -\gamma\boldsymbol{\beta} & \underline{1} - \hat{\boldsymbol{\beta}} \otimes \hat{\boldsymbol{\beta}}^T + \gamma\hat{\boldsymbol{\beta}} \otimes \hat{\boldsymbol{\beta}}^T \end{pmatrix} \tag{4.41}$$

yields with $\boldsymbol{\beta} = \boldsymbol{v}/c$ the electric and magnetic fields in the (ct', x', y', z') frame,

$$\boldsymbol{E}'(\boldsymbol{x}', t') = \gamma(\boldsymbol{E}(\boldsymbol{x}, t) + \boldsymbol{v} \times \boldsymbol{B}(\boldsymbol{x}, t)) - (\gamma - 1)\hat{\boldsymbol{\beta}}\left(\hat{\boldsymbol{\beta}} \cdot \boldsymbol{E}(\boldsymbol{x}, t)\right) \tag{4.42}$$

$$= \gamma(\boldsymbol{E}(\boldsymbol{x}, t) + \boldsymbol{v} \times \boldsymbol{B}(\boldsymbol{x}, t)) - \frac{\gamma^2}{(\gamma + 1)c^2}\boldsymbol{v}(\boldsymbol{v} \cdot \boldsymbol{E}(\boldsymbol{x}, t)), \tag{4.43}$$

$$\boldsymbol{B}'(\boldsymbol{x}', t') = \gamma\left(\boldsymbol{B}(\boldsymbol{x}, t) - \frac{1}{c^2}\boldsymbol{v} \times \boldsymbol{E}(\boldsymbol{x}, t)\right) - (\gamma - 1)\hat{\boldsymbol{\beta}}\left(\hat{\boldsymbol{\beta}} \cdot \boldsymbol{B}(\boldsymbol{x}, t)\right) \tag{4.44}$$

$$= \gamma\left(\boldsymbol{B}(\boldsymbol{x}, t) - \frac{1}{c^2}\boldsymbol{v} \times \boldsymbol{E}(\boldsymbol{x}, t)\right) - \frac{\gamma^2}{(\gamma + 1)c^2}\boldsymbol{v}(\boldsymbol{v} \cdot \boldsymbol{B}(\boldsymbol{x}, t)). \tag{4.45}$$

We can also express this in terms of the field strength components parallel and perpendicular to the relative velocity $\boldsymbol{v}$ between the two observers,

$$\boldsymbol{E}'_{\|}(\boldsymbol{x}', t') = \boldsymbol{E}_{\|}(\boldsymbol{x}, t), \quad \boldsymbol{B}'_{\|}(\boldsymbol{x}', t') = \boldsymbol{B}_{\|}(\boldsymbol{x}, t), \tag{4.46}$$

and

$$\boldsymbol{E}'_{\perp}(\boldsymbol{x}', t') = \gamma(\boldsymbol{E}_{\perp}(\boldsymbol{x}, t) + \boldsymbol{v} \times \boldsymbol{B}_{\perp}(\boldsymbol{x}, t)), \tag{4.47}$$

$$\boldsymbol{B}'_{\perp}(\boldsymbol{x}', t') = \gamma\left(\boldsymbol{B}_{\perp}(\boldsymbol{x}, t) - \frac{1}{c^2}\boldsymbol{v} \times \boldsymbol{E}_{\perp}(\boldsymbol{x}, t)\right). \tag{4.48}$$

These equations show that electric and magnetic fields mix under Lorentz transformations. The distinction between electric and magnetic fields therefore depends on the reference frame.

We can confirm through explicit substitutions of $F^{\mu\nu}$ and j^{ν} that the equations

$$\partial_\mu F^{\mu\nu} = -\mu_0 j^{\nu} \tag{4.49}$$

are the inhomogeneous Maxwell's equations, i.e. the differential forms of Coulomb's law and Ampère's law, respectively,

$$\nabla \cdot \boldsymbol{E} = \frac{1}{\epsilon_0}\varrho, \quad \nabla \times \boldsymbol{B} - \frac{1}{c^2}\frac{\partial}{\partial t}\boldsymbol{E} = \mu_0 \boldsymbol{j}, \tag{4.50}$$

while the identities

$$\epsilon^{\kappa\lambda\mu\nu}\partial_\lambda F_{\mu\nu} = 2\epsilon^{\kappa\lambda\mu\nu}\partial_\lambda\partial_\mu A_\nu \equiv 0 \tag{4.51}$$

(with the four-dimensional ϵ-tensor, $\epsilon^{0123} = -1$) are the homogeneous Maxwell's equations, i.e. Gauss' law of absence of magnetic monopoles and Faraday's law of induction, respectively,

$$\nabla \cdot \boldsymbol{B} = 0, \quad \nabla \times \boldsymbol{E} + \frac{\partial}{\partial t}\boldsymbol{B} = \boldsymbol{0}. \tag{4.52}$$

Equation (4.51) shows that these equations can be written in terms of the *dual field strength tensor*

$$\tilde{F}^{\mu\nu} = \frac{1}{2}\epsilon^{\mu\nu\alpha\beta}F_{\alpha\beta} = \begin{pmatrix} 0 & -B_1 & -B_2 & -B_3 \\ B_1 & 0 & E_3/c & -E_2/c \\ B_2 & -E_3/c & 0 & E_1/c \\ B_3 & E_2/c & -E_1/c & 0 \end{pmatrix} \tag{4.53}$$

in the form

$$\partial_\mu \tilde{F}^{\mu\nu} = 0. \tag{4.54}$$

The gauge transformation

$$A_\mu(x) \to A'_\mu(x) = A_\mu(x) + \partial_\mu f(x) \tag{4.55}$$

apparently leaves the field strength tensor $F_{\mu\nu}$ (4.37) invariant. In conventional terms equation (4.55) is

$$\Phi'(x) = \Phi(x) - \dot{f}(x), \quad \boldsymbol{A}'(x) = \boldsymbol{A}(x) + \nabla f(x). \tag{4.56}$$

We have expressed Maxwell's equations as equations between 4-vectors,

$$\partial_\mu F^{\mu\nu} = -\mu_0 j^\nu, \quad \partial_\mu \tilde{F}^{\mu\nu} = 0, \tag{4.57}$$

and this demonstrates that *Maxwell's equations hold in this form for every inertial observer*. Maxwell's equations have exactly the same form in every inertial reference frame. This is the *form invariance* (or simply 'invariance' or 'covariance') of Maxwell's equations under Lorentz transformations[5]. Recall that we could also write charge conservation in manifestly Lorentz invariant form (4.34), but to complete the manifestly covariant formulation of electrodynamics, we also need to write the Lorentz force law in 4-vector notation.

Electromagnetic forces in STR

The non-relativistic equation of motion for a particle with electric charge q in electromagnetic fields $\boldsymbol{E}$ and $\boldsymbol{B}$ contains the Lorentz force $\boldsymbol{F} = q\boldsymbol{E} + q\boldsymbol{v} \times \boldsymbol{B}$,

$$\frac{d}{dt}\boldsymbol{p}(t) = m\ddot{\boldsymbol{x}}(t) = q\boldsymbol{E}(\boldsymbol{x}(t), t) + q\dot{\boldsymbol{x}}(t) \times \boldsymbol{B}(\boldsymbol{x}(t), t), \tag{4.58}$$

and the corresponding energy balance equation is

$$\frac{dE}{dt} = \frac{d}{dt}\frac{m}{2}\dot{\boldsymbol{x}}^2(t) = q\dot{\boldsymbol{x}}(t) \cdot \boldsymbol{E}(\boldsymbol{x}(t), t). \tag{4.59}$$

Expressing $\boldsymbol{F}$ in terms of the field strength tensor (4.37) will help us to determine the relativistic generalization of the equations (4.58) and (4.59). The electric and magnetic field components are

$$E^i = cF^i{}_0 = F^i{}_0\frac{dx^0}{dt}, \quad \varepsilon^i{}_{jk}B^k = F^i{}_j. \tag{4.60}$$

The latter equation implies in particular $(\boldsymbol{v} \times \boldsymbol{B})^i = \varepsilon^i{}_{jk}v^j B^k = F^i{}_j v^j$, and therefore we can write the components of the Lorentz force in the form

$$F^i = qE^i + q\varepsilon^i{}_{jk}v^j B^k = qF^i{}_0\frac{dx^0}{dt} + qF^i{}_j\frac{dx^j}{dt} = qF^i{}_\nu\frac{dx^\nu}{dt}. \tag{4.61}$$

We could identify this with the spatial part of a 4-vector if we would not calculate the derivative with respect to the lab time t, but with respect to the eigentime τ of the charged particle. This yields with $d/d\tau = \gamma d/dt$ the force components

[5] Although technically the notion of covariance refers to a particular tensorial transformation behavior under linear transformations (2.16) and (3.42), in physics *form invariance* of equations is usually simply denoted as *covariance* of those equations, because form invariance is a direct consequence of the same transformation behavior of both sides of an equation under coordinate transformations.

$$f^i = \gamma F^i = qF^i{}_\nu \frac{dx^\nu}{d\tau}, \tag{4.62}$$

and the corresponding time component follows as

$$f^0 = qF^0{}_i \frac{dx^i}{d\tau} = q\gamma \frac{1}{c} E_i \frac{dx^i}{dt} = \gamma q \boldsymbol{\beta} \cdot \boldsymbol{E}. \tag{4.63}$$

This yields the electromagnetic force 4-vector

$$f^\mu = qF^\mu{}_\nu \frac{dx^\nu}{d\tau} = (\gamma q \boldsymbol{\beta} \cdot \boldsymbol{E}, \gamma q(\boldsymbol{E} + \boldsymbol{v} \times \boldsymbol{B})). \tag{4.64}$$

The equation of motion of the charged particle in 4-vector notation can therefore be written as $md^2x^\mu/d\tau^2 = qF^\mu{}_\nu dx^\nu/d\tau$, or

$$\frac{d}{d\tau} p^\mu(\tau) = m\ddot{x}^\mu(\tau) = qF^\mu{}_\nu(x(\tau))\dot{x}^\nu(\tau). \tag{4.65}$$

The time component of these equations yields after rescaling with c/γ again the energy balance equation,

$$\frac{dE}{dt} = q\boldsymbol{v} \cdot \boldsymbol{E}, \tag{4.66}$$

and the spatial part is after rescaling with γ^{-1}:

$$\frac{d}{dt}\boldsymbol{p} = q(\boldsymbol{E} + \boldsymbol{v} \times \boldsymbol{B}). \tag{4.67}$$

In conclusion we note that the only changes in the equations (4.67) and (4.66) with respect to the non-relativistic equations (4.58) and (4.59) are the velocity dependences of E and $\boldsymbol{p}$:

$$\boldsymbol{p} = \frac{m\boldsymbol{v}}{\sqrt{1 - (v^2/c^2)}}, \quad E = \frac{mc^2}{\sqrt{1 - (v^2/c^2)}}. \tag{4.68}$$

The set of equations (4.66) and (4.67) is completely equivalent to the set of equations (4.65). The advantage of the formulation (4.65) is the *manifest covariance* of these equations, since linearly transforming equations between 4-vectors *obviously* must hold in every inertial frame. Contrary to this, covariance is *hidden* in the equations (4.66) and (4.67), but since they are equivalent to the manifestly covariant equations (4.65), they also must hold in every inertial frame. Practical calculations of particle paths in electromagnetic fields usually prefer the use of the lab time t and therefore use the equations (4.66) and (4.67).

4.3 Action principles for relativistic particles

Action principles are extremely powerful tools for the description of physical systems, and relativistic systems are no exception to this rule. The easiest access to learn the basics of action principles is through Lagrangian mechanics[6]. However, we will also develop the basics of Lagrangian mechanics along the way.

We will often have to deal with ∇-operators which refer to coordinate sets in different spaces, and also with ∇-operators in velocity space. Therefore it is much more convenient to write ∇-operators in the following form,

$$\frac{\partial}{\partial \boldsymbol{x}} \equiv \nabla_{x} = \boldsymbol{e}^i \frac{\partial}{\partial x^i}, \quad \frac{\partial}{\partial \boldsymbol{x}'} \equiv \nabla_{x'} = \boldsymbol{e}^i \frac{\partial}{\partial x'^i}, \quad \frac{\partial}{\partial \dot{\boldsymbol{x}}} \equiv \nabla_{\dot{x}} = \boldsymbol{e}^i \frac{\partial}{\partial \dot{x}^i}. \tag{4.69}$$

Action principle for a relativistic charged particle

Action principles in mechanics entail that the equations of motions of a particle with trajectory $\boldsymbol{x}(t)$ can be derived from the requirement that a certain action integral

$$S[\boldsymbol{x}(t)] = \int_{t_0}^{t_1} dt\; L(\boldsymbol{x}(t), \dot{\boldsymbol{x}}(t), t) \tag{4.70}$$

does not change under arbitrary first-order perturbations of the path $\boldsymbol{x}(t) \to \boldsymbol{x}(t) + \delta\boldsymbol{x}(t)$, with fixed endpoints $\boldsymbol{x}(t_0)$ and $\boldsymbol{x}(t_1)$,

$$\delta\boldsymbol{x}(t_0) = 0, \quad \delta\boldsymbol{x}(t_1) = 0. \tag{4.71}$$

This statement is *Hamilton's principle of stationary action*. The integrand $L(\boldsymbol{x}(t), \dot{\boldsymbol{x}}(t), t)$ is the *Lagrange function* for the particle, and the notation $S[\boldsymbol{x}(t)]$ indicates that S depends on the whole path $\boldsymbol{x}(t)$ from the initial point $\boldsymbol{x}(t_0)$ to the final point $\boldsymbol{x}(t_1)$.

To see how the requirement of unchanged action implies equations of motion for the particle, we calculate the first-order change of S under the path variation $\boldsymbol{x}(t) \to \boldsymbol{x}(t) + \delta\boldsymbol{x}(t)$,

$$\begin{aligned} \delta S[\boldsymbol{x}(t)] &= S[\boldsymbol{x}(t) + \delta\boldsymbol{x}(t)] - S[\boldsymbol{x}(t)] \\ &= \int_{t_0}^{t_1} dt\; [L(\boldsymbol{x}(t) + \delta\boldsymbol{x}(t), \dot{\boldsymbol{x}}(t) + \delta\dot{\boldsymbol{x}}(t), t) - L(\boldsymbol{x}(t), \dot{\boldsymbol{x}}(t), t)]. \end{aligned} \tag{4.72}$$

First-order expansion of the Lagrange function yields

$$\delta S[\boldsymbol{x}(t)] = \int_{t_0}^{t_1} dt \left(\delta\boldsymbol{x}(t) \cdot \frac{\partial L(\boldsymbol{x}(t), \dot{\boldsymbol{x}}(t), t)}{\partial \boldsymbol{x}(t)} + \delta\dot{\boldsymbol{x}}(t) \cdot \frac{\partial L(\boldsymbol{x}(t), \dot{\boldsymbol{x}}(t), t)}{\partial \dot{\boldsymbol{x}}(t)} \right),$$

and integration by parts in the second term yields

$$\delta S[\boldsymbol{x}(t)] = \int_{t_0}^{t_1} dt\; \delta\boldsymbol{x}(t) \cdot \left(\frac{\partial L(\boldsymbol{x}(t), \dot{\boldsymbol{x}}(t), t)}{\partial \boldsymbol{x}(t)} - \frac{d}{dt} \frac{\partial L(\boldsymbol{x}(t), \dot{\boldsymbol{x}}(t), t)}{\partial \dot{\boldsymbol{x}}(t)} \right),$$

[6] Readers who are not familiar with Lagrangian mechanics can find a nutshell introduction in appendix A in Dick R 2016 *Advanced Quantum Mechanics: Materials and Photons* 2nd edn (Cham: Springer).

where we used that fixation of the endpoints (4.71) eliminates the boundary terms at t_0 and t_1. Since Hamilton's principle requires $\delta S[\boldsymbol{x}(t)] = 0$ for every perturbation $\delta\boldsymbol{x}(t)$ along the path (with fixed endpoints), we find the differential equation for the path $\boldsymbol{x}(t)$,

$$\frac{d}{dt}\frac{\partial L(\boldsymbol{x}(t), \dot{\boldsymbol{x}}(t), t)}{\partial \dot{\boldsymbol{x}}(t)} = \frac{\partial L(\boldsymbol{x}(t), \dot{\boldsymbol{x}}(t), t)}{\partial \boldsymbol{x}(t)}. \tag{4.73}$$

This equation is the *Euler–Lagrange equation*.

Indeed, the equation of motion (4.67) for a charged particle in the presence of electromagnetic fields (4.36) with potentials $A_\mu(x) = \{-\Phi(x)/c, \boldsymbol{A}(x)\}$ can be derived from the following action,

$$\begin{aligned} S[x(t)] &= \int dt\ L(\boldsymbol{x}(t), \dot{\boldsymbol{x}}(t)) = -mc\int dt\sqrt{c^2 - \dot{\boldsymbol{x}}^2(t)} + q\int dt\ \dot{x}^\mu(t)A_\mu(x(t)) \\ &= -mc\int dt\sqrt{c^2 - \dot{\boldsymbol{x}}^2(t)} + q\int dt\ [\dot{\boldsymbol{x}}(t)\cdot\boldsymbol{A}(\boldsymbol{x}(t), t) - \Phi(\boldsymbol{x}(t), t)]. \end{aligned} \tag{4.74}$$

We have

$$\frac{\partial L(\boldsymbol{x}, \dot{\boldsymbol{x}})}{\partial \dot{\boldsymbol{x}}} = mc\frac{\dot{\boldsymbol{x}}}{\sqrt{c^2 - \dot{\boldsymbol{x}}^2}} + q\boldsymbol{A}, \tag{4.75}$$

and the Euler–Lagrange equation (4.73) yields just equation (4.67). In the derivation you have to use

$$\frac{d}{dt}\boldsymbol{A}(\boldsymbol{x}(t), t) = \frac{\partial}{\partial t}\boldsymbol{A}(\boldsymbol{x}, t) + \dot{\boldsymbol{x}}\cdot\frac{\partial}{\partial \boldsymbol{x}}\boldsymbol{A}(\boldsymbol{x}, t) \tag{4.76}$$

and

$$\dot{\boldsymbol{x}}\times\boldsymbol{B} = \dot{\boldsymbol{x}}\times(\nabla\times\boldsymbol{A}) = \boldsymbol{e}_i\left(\dot{\boldsymbol{x}}\cdot\frac{\partial}{\partial x^i}\boldsymbol{A} - \dot{\boldsymbol{x}}\cdot\frac{\partial}{\partial \boldsymbol{x}}A_i\right). \tag{4.77}$$

Action principle in terms of the eigentime

The action (4.74) for vanishing charge or vanishing electromagnetic fields,

$$S = -mc\int dt\sqrt{c^2 - \dot{\boldsymbol{x}}^2(t)} \tag{4.78}$$

is the action for a free particle in terms of the lab time t. However, using

$$c^2 d\tau^2 = -\eta_{\mu\nu}dx^\mu dx^\nu = (c^2 - \dot{\boldsymbol{x}}^2(t))dt^2 \tag{4.79}$$

we can also write this as

$$S = -mc^2\int d\tau, \tag{4.80}$$

i.e. the free part of the action is just $-mc^2$ times the integrated eigentime of the moving particle, or $-mc$ times the length of the world line of the moving particle.

Furthermore, the requirement $\delta S = 0$ is of course independent of the parametrization of the integrals in S. We can get the manifestly covariant form (4.65) of the equations of motion by not singling out the lab time $t = x^0/c$ as the time parameter along the particle's world line, but by using the invariant eigentime of the moving particle instead,

$$\begin{aligned} S &= -mc^2 \int d\tau + q \int d\tau\, \dot{x}^\mu(\tau) A_\mu(x(\tau)) \\ &= -mc \int d\tau \sqrt{-\eta_{\mu\nu}\dot{x}^\mu(\tau)\dot{x}^\nu(\tau)} + q \int d\tau\, \dot{x}^\mu(\tau) A_\mu(x(\tau)). \end{aligned} \tag{4.81}$$

We have

$$\frac{\partial L}{\partial \dot{x}^\mu(\tau)} = mc\frac{\dot{x}_\mu(\tau)}{\sqrt{-\dot{x}^2(\tau)}} + qA_\mu(x(\tau)) = m\dot{x}_\mu(\tau) + qA_\mu(x(\tau)), \tag{4.82}$$

and the Euler–Lagrange equation yields indeed

$$\frac{d}{d\tau}\frac{\partial L}{\partial \dot{x}^\mu} - \frac{\partial L}{\partial x^\mu} = m\ddot{x}_\mu + q\dot{x}^\nu\partial_\nu A_\mu - q\dot{x}^\nu\partial_\mu A_\nu = m\ddot{x}_\mu - qF_{\mu\nu}\dot{x}^\nu = 0. \tag{4.83}$$

In equation (4.81) I have used

$$c^2 d\tau^2 = -\eta_{\mu\nu} dx^\mu dx^\nu = -\eta_{\mu\nu}\frac{dx^\mu}{d\tau}\frac{dx^\nu}{d\tau}d\tau^2, \tag{4.84}$$

which implies $-\eta_{\mu\nu}\dot{x}^\mu(\tau)\dot{x}^\nu(\tau) = c^2$.

4.4 Current densities and stress–energy tensors

If a conserved quantity has volume density ϱ, then the amount Q of the conserved quantity in a volume V is

$$Q = \int_V d^3\boldsymbol{x}\, \varrho. \tag{4.85}$$

This applies in particular to electric charge, but also to all other conserved quantities. Due to the familiar example of electric charge, it is customary to denote Q as a charge, no matter whether we are dealing with electric charge or another conserved quantity.

We also assume that V is a constant fixed volume. Since the quantity with density ϱ is conserved in the Universe, the charge Q in the fixed volume V can only change if it enters or escapes through the boundary ∂V of V,

$$\frac{dQ}{dt} = \frac{d}{dt}\int_V d^3\boldsymbol{x}\, \varrho = -\oint_{\partial V} d^2S\, \hat{\boldsymbol{n}}_S \cdot \boldsymbol{j}, \tag{4.86}$$

where $\hat{\boldsymbol{n}}_S$ is the outwards directed normal vector on the boundary ∂V with surface element d^2S. The vector $\boldsymbol{j}$ is the current density of the charge, i.e. the magnitude $|\boldsymbol{j}|$ is charge per time and per area A flowing in the direction of $\boldsymbol{j}$, where the area is measured perpendicular to the direction of $\boldsymbol{j}$. According to the Gauss theorem, equation (4.86) implies

$$\int_V d^3\boldsymbol{x}\left(\frac{\partial}{\partial t}\varrho + \nabla \cdot \boldsymbol{j}\right) = 0, \tag{4.87}$$

and since this must hold for every fixed volume V, no matter how small, equation (4.86) is equivalent to the *local conservation law*

$$\frac{\partial}{\partial t}\varrho + \nabla \cdot \boldsymbol{j} = \partial_\mu j^\mu = 0, \quad j^\mu = (\varrho c, \boldsymbol{j}). \tag{4.88}$$

We have encountered this equation already in the particular case of electric charges (4.34), and the previous reasoning tells us that an equation of this form has to hold for every conserved quantity.

Since the charge density of a particle with charge q on a trajectory $\boldsymbol{x}(t)$ is $\varrho(\boldsymbol{x}, t) = q\delta^3(\boldsymbol{x} - \boldsymbol{x}(t))$, the 4-current density of the particle is $j^\mu = \{\varrho c, \varrho \boldsymbol{v}\} = \varrho U^\mu/\gamma$:

$$\begin{aligned} j^0(\boldsymbol{x}, t) = \varrho(\boldsymbol{x}, t)c &= qc\delta^3(\boldsymbol{x} - \boldsymbol{x}(t)) \\ &= q\dot{x}^0(\tau)\gamma^{-1}(\tau)\delta^3(\boldsymbol{x} - \boldsymbol{x}(\tau)), \end{aligned} \tag{4.89}$$

$$\begin{aligned} \boldsymbol{j}(\boldsymbol{x}, t) = \varrho(\boldsymbol{x}, t)\boldsymbol{v}(t) &= q\boldsymbol{v}(t)\delta^3(\boldsymbol{x} - \boldsymbol{x}(t)) \\ &= q\dot{\boldsymbol{x}}(\tau)\gamma^{-1}(\tau)\delta^3(\boldsymbol{x} - \boldsymbol{x}(\tau)), \end{aligned} \tag{4.90}$$

where from equation (4.11), $\gamma(\tau) = \dot{x}^0(\tau)/c = \sqrt{1 + (\dot{\boldsymbol{x}}^2(\tau)/c^2)}$.

We can easily confirm charge conservation using the t-dependent representations of the components of the current density,

$$\begin{aligned} \partial_\mu j^\mu(\boldsymbol{x}, t) &= q\left(\frac{\partial}{\partial t} + \boldsymbol{v}(t) \cdot \frac{\partial}{\partial \boldsymbol{x}}\right)\delta^3(\boldsymbol{x} - \boldsymbol{x}(t)) \\ &= q\left(\frac{\partial}{\partial t} - \dot{\boldsymbol{x}}(t) \cdot \frac{\partial}{\partial \boldsymbol{x}(t)}\right)\delta^3(\boldsymbol{x} - \boldsymbol{x}(t)) \equiv 0. \end{aligned} \tag{4.91}$$

Equations (4.85)–(4.88) apply to scalar conserved quantities like the electric charge. But what if the conserved quantity is a 4-vector of conserved charges, like e.g. the 4-momentum p^μ of a system in a flat spacetime? Equation (4.85) is then replaced by an equation containing a 4-momentum density $\mathcal{P}^\mu$,

$$p^\mu = \int_V d^3\boldsymbol{x}\, \mathcal{P}^\mu, \tag{4.92}$$

and because the momentum in the fixed volume V can only change due to flow through the boundaries,

$$\frac{d}{dt}p^\mu = \frac{d}{dt}\int_V d^3\boldsymbol{x}\ \mathcal{P}^\mu = -\oint_{\partial V} d^2S\ \boldsymbol{T}^\mu \cdot \hat{\boldsymbol{n}}_S, \tag{4.93}$$

we find a local conservation law

$$\frac{\partial}{\partial t}\mathcal{P}^\mu + \nabla \cdot \boldsymbol{T}^\mu = c\partial_0\mathcal{P}^\mu + \nabla \cdot \boldsymbol{T}^\mu = \partial_\nu T^{\mu\nu} = 0. \tag{4.94}$$

The tensor components with a timelike index as the second index apparently yield the 4-momentum density,

$$\frac{1}{c}T^{\mu 0} = \mathcal{P}^\mu, \tag{4.95}$$

i.e. $\mathcal{P}^i = T^{i0}/c$ is the density of momentum component ^{pi} in $\boldsymbol{e}_i$ direction, while due to $E = cp^0$ the tensor component $T^{00} = c\mathcal{P}^0$ is the energy density. The tensor components with a spacelike index $T^{\mu i}$ as the second index yield the current density $\boldsymbol{T}^\mu = T^{\mu i}\boldsymbol{e}_i$ of the momentum component p^μ in the spatial direction $\boldsymbol{e}_i$, i.e. $T^{\mu i}$ is the amount of p^μ transported per time and per area in the direction $\boldsymbol{e}_i$.

Equation (4.93) also implies a dual interpretation of T^{ij} besides being the current density of momentum ^{pi} in the direction $\boldsymbol{e}_j$. Repeating the spatial components of the equation,

$$\frac{d}{dt}p^i = -\oint_{\partial V} d^2S\ \boldsymbol{T}^i \cdot \hat{\boldsymbol{n}}_S, \tag{4.96}$$

shows that $-T^{ij}$ is also the force in direction $\boldsymbol{e}_i$ per unit of surface area with surface normal $\boldsymbol{e}_j$, exerted **on** the system which carries the momentum $\boldsymbol{p}$. For example $-T^{12}$ is the force in x-direction per unit of area in the (x, z)-plane. *Action = reaction* then also implies that T^{ij} is the force in direction $\boldsymbol{e}_i$ per unit of surface area with surface normal $\boldsymbol{e}_j$, exerted **by** the system which carries the momentum $\boldsymbol{p}$.

The coefficients $T_{\mu\nu}$ form a symmetric tensor[7], $T_{\mu\nu} = T_{\nu\mu}$. This tensor is denoted as the *stress–energy tensor* or *energy–momentum tensor*.

The stress–energy tensor of electromagnetic fields

Energy conservation for electromagnetic fields in the absence of charges reads in an inertial frame[8]

[7] More specifically, the stress–energy tensor is symmetric in relativistic theories, but it can be non-symmetric in the sense $T^{i0} \neq T^{0i}$ in nonrelativistic theories which neglect rest mass terms. Furthermore, we can also have $T_{ij} \neq T_{ji}$ in systems which are not rotationally symmetric.

[8] If you would like to know how these balance equations for energy and momentum of electromagnetic fields are derived, you can find the derivation of the electromagnetic stress–energy tensor (4.101) in the field theory generalization of Lagrangian mechanics in chapters 16 and 18 in Dick R 2016 *Advanced Quantum Mechanics: Materials and Photons*, 2nd edn (Cham: Springer). Chapter 6 in Jackson J D 1999 *Classical Electrodynamics*, 3rd edn (New York: Wiley), shows the derivation from the equations of motion of electrodynamics. However, for the purposes of this course our concern is not the derivation of the energy–momentum densities and currents of electromagnetic fields, but their applications in relativity theory.

$$\frac{\partial}{\partial t}\left(\frac{\epsilon_0}{2}\boldsymbol{E}^2 + \frac{1}{2\mu_0}\boldsymbol{B}^2\right) + \nabla \cdot \frac{1}{\mu_0}(\boldsymbol{E} \times \boldsymbol{B}), = 0 \tag{4.97}$$

or expressed in terms of the electromagnetic field strength tensor (4.37),

$$\frac{\partial}{\partial t}\left(\frac{1}{2\mu_0}F^0{}_iF^{0i} + \frac{1}{4\mu_0}F_{ij}F^{ij}\right) + \partial_i\frac{c}{\mu_0}\left(F^{0j}F^i{}_j\right) = 0. \tag{4.98}$$

Momentum conservation for electromagnetic fields in the absence of charges reads

$$\begin{aligned} &\frac{\partial}{\partial t}\epsilon_0(\boldsymbol{E} \times \boldsymbol{B}) + \nabla\left(\frac{\epsilon_0}{2}\boldsymbol{E}^2 + \frac{1}{2\mu_0}\boldsymbol{B}^2\right) \\ &- \nabla \cdot \left(\epsilon_0\boldsymbol{E} \otimes \boldsymbol{E}^T + \frac{1}{\mu_0}\boldsymbol{B} \otimes \boldsymbol{B}^T\right) = \boldsymbol{0}, \end{aligned} \tag{4.99}$$

or in terms of the field strength tensor:

$$\begin{aligned} &\frac{\partial}{\partial t}\epsilon_0 c\left(F^{0j}F^i{}_j\right) + \partial_j\left(-\delta^{ij}\frac{1}{2\mu_0}F_{0k}F^{0k} - \delta^{ij}\frac{1}{4\mu_0}F_{kl}F^{kl}\right. \\ &\left. + \frac{1}{\mu_0}F^i{}_0F^{j0} + \frac{1}{\mu_0}F^i{}_kF^{jk}\right) = 0. \end{aligned} \tag{4.100}$$

The conservation laws (4.98) and (4.100) for energy and momentum can also be written in the form (4.94), $\partial_\nu T^{\mu\nu} = 0$, with the stress–energy tensor of the electromagnetic field

$$T^{\mu\nu} = \frac{1}{\mu_0}\left(F^{\mu\lambda}F^\nu{}_\lambda - \frac{1}{4}\eta^{\mu\nu}F_{\kappa\lambda}F^{\kappa\lambda}\right), \tag{4.101}$$

or in explicit matrix form,

$$\begin{aligned} \underline{T} = \{T^{\mu\nu}\} &= \begin{pmatrix} \mathcal{H} & \boldsymbol{S}/c \\ c\boldsymbol{\mathcal{P}} & -\underline{\mathcal{M}} \end{pmatrix} \\ &= \begin{pmatrix} \frac{\epsilon_0}{2}\boldsymbol{E}^2 + \frac{1}{2\mu_0}\boldsymbol{B}^2 & c\epsilon_0(\boldsymbol{E} \times \boldsymbol{B})^T \\ c\epsilon_0(\boldsymbol{E} \times \boldsymbol{B}) & \left(\frac{\epsilon_0}{2}\boldsymbol{E}^2 + \frac{1}{2\mu_0}\boldsymbol{B}^2\right)\underline{\boldsymbol{1}} - \epsilon_0\boldsymbol{E} \otimes \boldsymbol{E}^T - \frac{1}{\mu_0}\boldsymbol{B} \otimes \boldsymbol{B}^T \end{pmatrix}. \end{aligned}$$

The conserved energy–momentum densities are $\mathcal{P}^\mu = \frac{1}{c}T^{\mu 0}$, i.e.

$$\mathcal{H} = c\mathcal{P}^0 = T^{00} = \frac{\epsilon_0}{2}\boldsymbol{E}^2 + \frac{1}{2\mu_0}\boldsymbol{B}^2 \tag{4.102}$$

is the energy density (here denoted by $\mathcal{H}$ because it appears as a Hamiltonian density if one quantizes the electromagnetic field to describe photons at the quantum level), and

$$\mathcal{P} = \frac{1}{\mu_0 c^2}\boldsymbol{E} \times \boldsymbol{B} = \epsilon_0 \boldsymbol{E} \times \boldsymbol{B} \tag{4.103}$$

is the momentum density. The vector

$$\boldsymbol{S} = \frac{1}{\mu_0}\boldsymbol{E} \times \boldsymbol{B} = c^2\mathcal{P} \tag{4.104}$$

is the Poynting vector for the energy current density in electromagnetic fields, and $\underline{\mathcal{M}}$ is Maxwell's stress tensor.

The stress–energy tensor of a perfect fluid

The stress–energy tensor in the local inertial rest frame of a perfect (i.e. isotropic, incompressible, non-viscous) fluid element has to be isotropic, i.e. invariant under rotations. Since the only vector invariant under rotations is the null-vector, the rest frame components must satisfy $T^{0i}|_r = T^{i0}|_r = 0$. Here the symbol $|_r$ reminds us that we are in the local rest frame of the fluid element.

Furthermore, the only class of matrices which are invariant under rotations are proportional to the unit matrix, and therefore we must have $T^{ij}|_r = p\delta^{ij}$, i.e. the only options for the stress–energy tensor in the local inertial rest frame are

$$T^{00}|_r = \varrho, \quad T^{0j}|_r = T^{j0}|_r = 0, \quad T^{ij}|_r = p\delta^{ij}. \tag{4.105}$$

We already know from the general discussion of stress–energy tensors that ϱ is the rest frame energy density in the fluid element. Furthermore, since equation (4.93) told us that T^{ij} is the force exerted by the fluid in direction $\boldsymbol{e}_i$, per area with normal vector $\boldsymbol{e}_j$, we can infer that p is the pressure in the fluid.

The Lorentz transformation from the momentary rest frame of the fluid element (the 'fluid frame') into the lab frame is

$$\begin{aligned}\Lambda &= \begin{pmatrix} \gamma & \gamma\boldsymbol{\beta}^T \\ \gamma\boldsymbol{\beta} & (\gamma - 1)\hat{\boldsymbol{v}} \otimes \hat{\boldsymbol{v}}^T + \underline{\mathbf{1}} \end{pmatrix} \\ &= \begin{pmatrix} u^0 & \boldsymbol{u}^T \\ \boldsymbol{u} & (u^0 - 1)\hat{\boldsymbol{u}} \otimes \hat{\boldsymbol{u}}^T + \underline{\mathbf{1}} \end{pmatrix},\end{aligned} \tag{4.106}$$

where the dimensionless 4-velocity u is defined as

$$u^\mu = \frac{1}{c}\frac{dx^\mu}{d\tau} = \frac{\gamma}{c}\frac{dx^\mu}{dt}, \tag{4.107}$$

$$u^0 = \gamma, \quad \boldsymbol{u} = \frac{\gamma}{c}\boldsymbol{v} = \gamma\boldsymbol{\beta}, \tag{4.108}$$

$$u^2 = \boldsymbol{u}^2 - (u^0)^2 = -1. \tag{4.109}$$

Note that the momentary fluid rest frame moves relative to the lab with the fluid velocity $\boldsymbol{v} = c\boldsymbol{\beta}$, and therefore the Lorentz transformation (4.106) from the fluid rest frame into the lab frame is inverse to the transformation from the lab frame into the fluid frame.

Application of the Lorentz transformation (4.106) yields the stress–energy tensor in the lab frame:

$$T^{00} = \Lambda^0{}_\mu \Lambda^0{}_\nu T^{\mu\nu}| = \varrho u^0 u^0 + p\boldsymbol{u}^2 = (\varrho + p)u^0 u^0 + p\eta^{00}, \tag{4.110}$$

$$T^{0j} = T^{j0} = \Lambda^0{}_\mu \Lambda^j{}_\nu T^{\mu\nu}| = \varrho u^0 u^j + p u^0 u^j, \tag{4.111}$$

$$\begin{aligned} T^{ij} = \Lambda^i{}_\mu \Lambda^j{}_\nu T^{\mu\nu}| &= \varrho u^i u^j + p(u^0 u^0 - 1)\frac{u^i u^j}{\boldsymbol{u}^2} + p\delta^{ij} \\ &= (\varrho + p)u^i u^j + p\eta^{ij}. \end{aligned} \tag{4.112}$$

In equation (4.112) we used the idempotence $\underline{\mathrm{P}}^2 = \underline{\mathrm{P}}$, the symmetry, and the mutual orthogonality of the projection operators $\hat{\boldsymbol{u}} \otimes \hat{\boldsymbol{u}}^T$ and $\underline{\mathbf{1}} - \hat{\boldsymbol{u}} \otimes \hat{\boldsymbol{u}}^T$,

$$\Lambda^i{}_k \Lambda^{jk} = [u^0 u^0 \hat{\boldsymbol{u}} \otimes \hat{\boldsymbol{u}}^T + \underline{\mathbf{1}} - \hat{\boldsymbol{u}} \otimes \hat{\boldsymbol{u}}^T]^{ij}. \tag{4.113}$$

The results for the different components of the stress–energy tensor of the perfect fluid can be summarized in the form

$$T^{\mu\nu} = (\varrho + p)u^\mu u^\nu + p\eta^{\mu\nu}. \tag{4.114}$$

The stress–energy tensor of a particle

We found the expression $p^\mu(\tau) = mU^\mu(\tau) = m\dot{x}^\mu(\tau)$ in equations (4.14) and (4.15) for the energy–momentum vector of a particle. The corresponding energy–momentum densities are then

$$\mathcal{P}^\mu = T^{\mu 0}/c = p^\mu(\tau)\delta^3(\boldsymbol{x} - \boldsymbol{x}(\tau)), \tag{4.115}$$

and if we substitute $1 = \dot{x}^0(\tau)/\gamma(\tau)c$ (e.g. from equation (4.11)), we find

$$T^{\mu 0} = \mathcal{P}^\mu c = m\dot{x}^\mu(\tau)\dot{x}^0(\tau)\gamma^{-1}(\tau)\delta^3(\boldsymbol{x} - \boldsymbol{x}(\tau)). \tag{4.116}$$

On the other hand, the current density components $T^{\mu i}$ describe the flow of energy–momentum density $\mathcal{P}^\mu$ in direction $\boldsymbol{e}_i$, and since the particles move with velocity components $v^i = U^i/\gamma$, we find

$$T^{\mu i} = \mathcal{P}^\mu v^i = m\dot{x}^\mu(\tau)\dot{x}^i(\tau)\gamma^{-1}(\tau)\delta^3(\boldsymbol{x} - \boldsymbol{x}(\tau)). \tag{4.117}$$

We can combine these results in the stress–energy tensor

$$\begin{aligned} T^{\mu\nu} &= p^\mu(\tau)\dot{x}^\nu(\tau)\gamma^{-1}(\tau)\delta^3(\boldsymbol{x} - \boldsymbol{x}(\tau)) \\ &= m\dot{x}^\mu(\tau)\dot{x}^\nu(\tau)\gamma^{-1}(\tau)\delta^3(\boldsymbol{x} - \boldsymbol{x}(\tau)), \end{aligned} \tag{4.118}$$

and we can relate this expression to the electric current density (4.89) and (4.90) for the particle in the form

$$T^{\mu\nu} = p^{\mu} j^{\nu}/q, \tag{4.119}$$

i.e. the conserved charge q is simply traded for the conserved charges p^{μ} in the comparison between the electric current density and the energy–momentum current densities of the particle.

We can easily verify the energy–momentum conservation laws in the absence of external forces: The free relativistic Newton equation $dp^{\mu}(\tau)/d\tau = 0$ implies

$$\partial_{\nu} p^{\mu} = \eta_{\nu}{}^{0}(\partial_0 \tau)(dp^{\mu}(\tau)/d\tau) = 0, \tag{4.120}$$

and therefore

$$\partial_{\nu} T^{\mu\nu} = p^{\mu} \partial_{\nu} j^{\nu}/q = 0, \tag{4.121}$$

cf equation (4.91).

The corresponding energy and momentum densities are given by the usual expressions, of course, if we switch back to lab time t as the time parameter,

$$\mathcal{H} = c\mathcal{P}^0 = T^{00} = \frac{mc^3}{\sqrt{c^2 - v^2}} \delta(\boldsymbol{x} - \boldsymbol{x}(t)) = E\delta^3(\boldsymbol{x} - \boldsymbol{x}(t)), \tag{4.122}$$

$$\mathcal{P} = \frac{1}{c} \boldsymbol{e}_i T^{i0} = \frac{mc\boldsymbol{v}}{\sqrt{c^2 - v^2}} \delta(\boldsymbol{x} - \boldsymbol{x}(t)) = \boldsymbol{p}\delta^3(\boldsymbol{x} - \boldsymbol{x}(t)). \tag{4.123}$$

In concluding, we note that with $d\tau = dt(\tau)/\gamma(\tau) = dx^0(\tau)/c\gamma(\tau)$, the result (4.118) can also be written in a manifestly Lorentz covariant form which does not *a priori* single out lab time,

$$\begin{aligned} T^{\mu\nu} &= mc \int d\tau\, \dot{x}^{\mu}(\tau) \dot{x}^{\nu}(\tau) \delta^4(x - x(\tau)) \\ &= c \int d\tau\, p^{\mu}(\tau) \dot{x}^{\nu}(\tau) \delta^4(x - x(\tau)). \end{aligned} \tag{4.124}$$

The total stress–energy tensor for a coupled system of charged particle and electromagnetic fields is

$$\begin{aligned} T_{\nu}{}^{\mu}(x) = &\frac{1}{\mu_0}\left(F_{\nu\rho}(x) F^{\mu\rho}(x) - \frac{1}{4}\eta_{\nu}{}^{\mu} F_{\rho\sigma}(x) F^{\rho\sigma}(x)\right) \\ &+ m\dot{x}_{\nu}(\tau)\dot{x}^{\mu}(\tau)\gamma^{-1}(\tau)\delta^3(\boldsymbol{x} - \boldsymbol{x}(\tau)) \\ = &\frac{1}{\mu_0}\left(F_{\nu\rho}(x) F^{\mu\rho}(x) - \frac{1}{4}\eta_{\nu}{}^{\mu} F_{\rho\sigma}(x) F^{\rho\sigma}(x)\right) + \frac{1}{q} p_{\nu}(\tau) j^{\mu}(x). \end{aligned} \tag{4.125}$$

The conservation law $\partial_{\mu} T^{\nu\mu} = 0$ can be verified from the Lorentz force law (4.65) and Maxwell's equations.

Chapter 5

Differential geometry: the kinematics of curved spacetime

We are now at a stage where we can move on to the next level, viz the generalization of relativity to curved spacetime. However, we have to familiarize ourselves with a few more notions from geometry before we can actually move into general relativity.

5.1 More geometry: surfaces in $\mathbb{R}^3$

We have to become familiar with the geometry of curved spaces, and the geometry of curved surfaces in an ambient flat Euclidean $\mathbb{R}^3$ is a good starting point to discuss many relevant notions. We will rediscover several notions which we encountered already in the discussion of curvilinear coordinates in flat spaces, and a few more.

In terms of local coordinate patches our surface can be described by

$$x^\mu = \{x^1, x^2\} \quad \Rightarrow \quad X^I(x) = \{X^1(x^1, x^2), X^2(x^1, x^2), X^3(x^1, x^2)\}, \tag{5.1}$$

or

$$x^\mu = \{x^1, x^2\} \quad \Rightarrow \quad \boldsymbol{r}(x) = X^I(x)\boldsymbol{e}_I, \tag{5.2}$$

where $\{\boldsymbol{e}_I\}$ is a Cartesian basis of $\mathbb{R}^3$. We denote the surface as $\mathcal{M}$. The point with three-dimensional coordinate vector $\boldsymbol{r}(x)$ is a point on $\mathcal{M}$, and the tangent plane at that point is denoted as $T_{\boldsymbol{r}(x)}\mathcal{M} \equiv T_x\mathcal{M}$.

The coordinate lines $x^2 = \text{const.}$ and $x^1 = \text{const.}$ in $\mathbb{R}^2$ are mapped onto corresponding coordinate lines on the embedded surface $\mathcal{M}$, and we can construct a basis of tangent vectors at each point $\boldsymbol{r}(x)$ of the coordinate patch $x \to \boldsymbol{r}(x)$ in terms of tangent vectors to the x^μ coordinate lines in the following way: Small coordinate changes δx^μ induce small shifts on the surface,

$$\delta\boldsymbol{r}(x) = \boldsymbol{r}(x + \delta x) - \boldsymbol{r}(x) = \delta x^\mu \partial_\mu \boldsymbol{r}(x) = \delta x^\mu \partial_\mu X^I(x)\boldsymbol{e}_I, \tag{5.3}$$

and this tells us that the vector

doi:10.1088/2053-2571/aaf173ch5

$$\boldsymbol{E}_\mu(x) = \partial_\mu \boldsymbol{r}(x) = \partial_\mu X^I(x)\boldsymbol{e}_I \tag{5.4}$$

is a tangent vector to the x^μ coordinate line on the surface $\mathcal{M}$ at the point $\boldsymbol{r}(x)$. Furthermore, the two tangent vectors $\boldsymbol{E}_1(x)$ and $\boldsymbol{E}_2(x)$ span the tangent vector space $T_x\mathcal{M}$ at the point[1] $\boldsymbol{r}(x)$, such that every tangent vector $\boldsymbol{v} \in T_x\mathcal{M}$ has an expansion $\boldsymbol{v} = v^\mu \boldsymbol{E}_\mu(x)$. A tangent vector field in a domain on $\mathcal{M}$ can also be expanded in this basis, $\boldsymbol{v}(x) = v^\mu(x)\boldsymbol{E}_\mu(x)$.

The distance squared between two nearby points $\boldsymbol{r}(x)$ and $\boldsymbol{r}(x + dx)$ on $\mathcal{M}$ is in leading order

$$\begin{aligned} ds^2 &= [\boldsymbol{r}(x + dx) - \boldsymbol{r}(x)]^2 = dx^\mu dx^\nu \partial_\mu X^I(x)\partial_\nu X^J(x)\boldsymbol{e}_I \cdot \boldsymbol{e}_J \\ &= \boldsymbol{E}_\mu(x) \cdot \boldsymbol{E}_\nu(x) dx^\mu dx^\nu = g_{\mu\nu}(x) dx^\mu dx^\nu. \end{aligned} \tag{5.5}$$

This defines the *induced metric* on the surface,

$$\begin{aligned} g_{\mu\nu}(x) &= \boldsymbol{E}_\mu(x) \cdot \boldsymbol{E}_\nu(x) = \partial_\mu X^I(x)\partial_\nu X^J(x)\boldsymbol{e}_I \cdot \boldsymbol{e}_J \\ &= \partial_\mu X^I(x)\partial_\nu X_I(x). \end{aligned} \tag{5.6}$$

The inverse metric $g^{\mu\nu}(x)$, $g^{\mu\nu}(x)g_{\nu\lambda}(x) = \delta^\mu{}_\lambda$, implies that the tangent vectors $\boldsymbol{E}^\mu(x) = g^{\mu\nu}(x)\boldsymbol{E}_\nu(x)$ form a dual basis of $T_x\mathcal{M}$,

$$\boldsymbol{E}^\mu(x) \cdot \boldsymbol{E}_\nu(x) = \delta^\mu{}_\nu, \tag{5.7}$$

and the inverse metric can be written as the scalar product of the dual basis vectors, $g^{\mu\nu}(x) = \boldsymbol{E}^\mu(x) \cdot \boldsymbol{E}^\nu(x)$.

A basis of the tangent space $T_x\mathbb{R}^3$ of $\mathbb{R}^3$ in the point $\boldsymbol{r}(x)$, which includes $T_x\mathcal{M}$ explicitly as a subspace, is given by the two tangent vectors (5.4) and the normal vector to the surface,

$$\begin{aligned} \boldsymbol{E}_\perp(x) &= \boldsymbol{E}_1(x) \times \boldsymbol{E}_2(x) = \partial_1 \boldsymbol{r} \times \partial_2 \boldsymbol{r} = \partial_1 X^I(x)\partial_2 X^J(x)\boldsymbol{e}_I \times \boldsymbol{e}_J \\ &= \partial_1 X^I(x)\partial_2 X^J(x)\epsilon_{IJK}\boldsymbol{e}^K. \end{aligned} \tag{5.8}$$

Furthermore, the projector onto $T_x\mathcal{M}$

$$\underline{\mathrm{P}}(x) = \boldsymbol{E}_\mu(x) \otimes \boldsymbol{E}^\mu(x), \quad \underline{\mathrm{P}}(x) \cdot \boldsymbol{E}_\mu(x) = \boldsymbol{E}_\mu(x), \quad \underline{\mathrm{P}}(x) \cdot \boldsymbol{E}_\perp(x) = 0, \tag{5.9}$$

projects every vector in $T_x\mathbb{R}^3$ onto its component in $T_x\mathcal{M}$.

5.2 Covariant derivatives and Christoffel symbols

The first step in calculating derivatives of vector fields $\boldsymbol{v}(x)$ in a curved space seems to require a comparison of the vector field in different points x and $x + \delta x$, before the limit $\delta x \to 0$ is taken. This is not straightforward, because $T_x\mathcal{M}$ and $T_{x+\delta x}\mathcal{M}$ generically are different vector spaces (different tangent planes). However, we can

[1] The surface generated by the mapping $x \to X(x)$ would be pinched in the point $X(x')$ if the two tangent vectors $\boldsymbol{E}_1(x')$ and $\boldsymbol{E}_2(x')$ are linearly dependent in the point $X(x')$. We assume that the surface is not pinched anywhere, and that the vectors $\boldsymbol{E}_1(x)$ and $\boldsymbol{E}_2(x)$ remain linearly independent everywhere.

circumvent this problem by first parallel transporting $\boldsymbol{v}(x) \in T_x\mathcal{M}$ to $x + \delta x$ in the ambient $\mathbb{R}^3$, and then project it down to $T_{x+\delta x}\mathcal{M}$ with the help of the projector $\underline{\mathrm{P}}(x + \delta x)$. The first-order expansion of the projector

$$\underline{\mathrm{P}}(x + \delta x) = \underline{\mathrm{P}}(x) + \delta x^\mu \partial_\mu [\boldsymbol{E}_\nu(x) \otimes \boldsymbol{E}^\nu(x)] \tag{5.10}$$

yields the parallel transported vector in first order:

$$\begin{aligned} \boldsymbol{v}(x, \delta x) &\equiv \underline{\mathrm{P}}(x + \delta x) \cdot \boldsymbol{v}(x) \\ &= \boldsymbol{v}(x) + \delta x^\mu \Big[v^\nu(x) \partial_\mu \boldsymbol{E}_\nu(x) - v^\rho(x) \boldsymbol{E}_\nu(x) \boldsymbol{E}^\nu(x) \cdot \partial_\mu \boldsymbol{E}_\rho(x) \Big], \end{aligned} \tag{5.11}$$

where we used $\partial_\mu \boldsymbol{E}^\nu(x) \cdot \boldsymbol{E}_\rho(x) = -\boldsymbol{E}^\nu(x) \cdot \partial_\mu \boldsymbol{E}_\rho(x)$ from equation (5.7).

Now we can define a new vector as the difference vector between the vector $\boldsymbol{v}(x + \delta x)$ and the vector $\boldsymbol{v}(x, \delta x)$:

$$\begin{aligned} \boldsymbol{v}(x + \delta x) - \boldsymbol{v}(x, \delta x) &= \delta x^\mu \boldsymbol{E}_\nu(x) \Big[\partial_\mu v^\nu(x) + v^\rho(x) \boldsymbol{E}^\nu(x) \cdot \partial_\mu \boldsymbol{E}_\rho(x) \Big] \\ &= \underline{\mathrm{P}}(x) \cdot \delta x^\mu \partial_\mu \boldsymbol{v}(x) \\ &= \delta x^\mu \boldsymbol{E}_\nu(x) \Big[\partial_\mu v^\nu(x) + \Gamma^\nu{}_{\rho\mu}(x) v^\rho(x) \Big] \\ &= \delta x^\mu \boldsymbol{E}_\nu(x) D_\mu v^\nu(x) \\ &= \delta x^\mu D_\mu \boldsymbol{v}(x) = D_{\delta x} \boldsymbol{v}(x). \end{aligned} \tag{5.12}$$

This defines the *covariant derivative* $D_{\delta x}\boldsymbol{v}(x)$ of the vector $\boldsymbol{v}(x)$ in the direction δx. We encounter again the Christoffel symbols[2]

$$\begin{aligned} \Gamma^\nu{}_{\rho\mu} &= \boldsymbol{E}^\nu \cdot \partial_\mu \boldsymbol{E}_\rho = \boldsymbol{E}^\nu \cdot \partial_\mu \partial_\rho \boldsymbol{r} = \boldsymbol{E}^\nu \cdot \partial_\rho \boldsymbol{E}_\mu \\ &= \frac{1}{2} g^{\nu\sigma} \left(\boldsymbol{E}_\sigma \cdot \partial_\mu \boldsymbol{E}_\rho + \boldsymbol{E}_\sigma \cdot \partial_\rho \boldsymbol{E}_\mu \right) \\ &= \frac{1}{2} g^{\nu\sigma} \Big[\partial_\mu \big(\boldsymbol{E}_\sigma \cdot \boldsymbol{E}_\rho \big) + \partial_\rho \big(\boldsymbol{E}_\sigma \cdot \boldsymbol{E}_\mu \big) - \partial_\sigma \big(\boldsymbol{E}_\rho \cdot \boldsymbol{E}_\mu \big) \Big] \\ &= \frac{1}{2} g^{\nu\sigma} \left(\partial_\mu g_{\sigma\rho} + \partial_\rho g_{\sigma\mu} - \partial_\sigma g_{\rho\mu} \right). \end{aligned} \tag{5.13}$$

On the other hand, if we expand the vector field in the dual basis, $\boldsymbol{v}(x) = v_\nu(x)\boldsymbol{E}^\nu(x)$, we find

$$\begin{aligned} D_{\delta x}\boldsymbol{v}(x) &= \delta x^\mu D_\mu \boldsymbol{v}(x) \equiv \boldsymbol{v}(x + \delta x) - \underline{\mathrm{P}}(x + \delta x) \cdot \boldsymbol{v}(x) \\ &= \delta x^\mu \boldsymbol{E}^\nu(x) \Big[\partial_\mu v_\nu(x) - v_\rho(x) \boldsymbol{E}^\rho(x) \cdot \partial_\mu \boldsymbol{E}_\nu(x) \Big] \\ &= \delta x^\mu \boldsymbol{E}^\nu(x) \Big[\partial_\mu v_\nu(x) - v_\rho(x) \Gamma^\rho{}_{\nu\mu}(x) \Big], \end{aligned} \tag{5.14}$$

i.e. the representations of the covariant derivative $D_\mu \boldsymbol{v}(x)$ in terms of the action on the contravariant or covariant vector components are

$$D_\mu v^\nu(x) = \partial_\mu v^\nu(x) + \Gamma^\nu{}_{\rho\mu}(x) v^\rho(x), \tag{5.15}$$

[2] Christoffel E B 1869 *Journal für die reine und angewandte Mathematik* **70** 46.

$$D_\mu v_\nu(x) = \partial_\mu v_\nu(x) - v_\rho(x)\Gamma^\rho{}_{\nu\mu}(x). \tag{5.16}$$

The covariant derivative of the vector field $\boldsymbol{v}(x)$ along the line $x = x(\xi)$ with parameter ξ is then

$$\begin{aligned} D_\xi \boldsymbol{v}(x) &= \frac{dx^\mu}{d\xi}\boldsymbol{E}_\nu(x) D_\mu v^\nu(x) \\ &= \frac{dx^\mu}{d\xi}[\partial_\mu v^\nu(x) + \Gamma^\nu{}_{\rho\mu}(x) v^\rho(x)]\boldsymbol{E}_\nu(x). \end{aligned} \tag{5.17}$$

We also note that the covariant derivatives of the tangent vectors $\boldsymbol{E}_\rho(x)$ are

$$\begin{aligned} D_{\delta x}\boldsymbol{E}_\rho(x) &= \boldsymbol{E}_\rho(x + \delta x) - \boldsymbol{E}_\rho(x, \delta x) = \delta x^\mu \boldsymbol{E}_\nu(x)\boldsymbol{E}^\nu(x) \cdot \partial_\mu \boldsymbol{E}_\rho(x) \\ &= \underline{\mathrm{P}}(x) \cdot \delta x^\mu \partial_\mu \boldsymbol{E}_\rho(x) = \delta x^\mu \boldsymbol{E}_\nu(x)\Gamma^\nu{}_{\rho\mu}(x). \end{aligned} \tag{5.18}$$

The *parallel transported* vector is

$$\begin{aligned} \boldsymbol{v}(x, \delta x) &= \boldsymbol{v}(x + \delta x) - D_{\delta x}\boldsymbol{v}(x) \\ &= \left[v^\nu(x) - \Gamma^\nu{}_{\rho\mu}(x) v^\rho(x)\delta x^\mu\right]\boldsymbol{E}_\nu(x + \delta x). \end{aligned} \tag{5.19}$$

Therefore, on a manifold which is not embedded in an ambient space, the definition of Christoffel symbols (e.g. from the metric as in equation (5.13)) can be interpreted as a prescription for parallel transport on a space.

Other possible definitions of $D_{\delta x}\boldsymbol{v}(x)$ from parallel transport

An expanded version of equation (5.12) also includes two other possible (but equivalent) definitions of $D_{\delta x}\boldsymbol{v}(x)$ from parallel transport. Projection of $\boldsymbol{v}(x - \delta x)$ into $T_x\mathcal{M}$ and comparison with $\boldsymbol{v}(x)$, or projection of $\boldsymbol{v}(x + \delta x)$ into $T_x\mathcal{M}$ and comparison with $\boldsymbol{v}(x)$ yield the same result for $D_{\delta x}\boldsymbol{v}(x)$ after first-order expansion in δx^μ,

$$\begin{aligned} \boldsymbol{v}(x + \delta x) - \boldsymbol{v}(x, \delta x) &= \underline{\mathrm{P}}(x) \cdot \delta x^\mu \partial_\mu \boldsymbol{v}(x) \\ &= \boldsymbol{v}(x) - \boldsymbol{v}(x - \delta x, \delta x) \\ &= \boldsymbol{v}(x) - \underline{\mathrm{P}}(x) \cdot \boldsymbol{v}(x - \delta x) \\ &= \boldsymbol{v}(x + \delta x, -\delta x) - \boldsymbol{v}(x) \\ &= \underline{\mathrm{P}}(x) \cdot \boldsymbol{v}(x + \delta x) - \boldsymbol{v}(x) \\ &= \delta x^\mu D_\mu \boldsymbol{v}(x) = D_{\delta x}\boldsymbol{v}(x). \end{aligned} \tag{5.20}$$

Parallel transport along a line

We have defined the parallel transport of a vector $\boldsymbol{v}(x) \in T_x\mathcal{M}$ to a vector $\boldsymbol{v}(x, \delta x) \in T_{x+\delta x}\mathcal{M}$ through equation (5.19). The Italian geometer Tullio Levi-Cività observed in 1917 that we can use this equation to start from the vector

$\mathbf{v}(x_0)$ in some arbitrary fixed point $X(x_0) \in \mathcal{M}$ and construct a new vector field $\mathbf{v}_{\|}(x)$ along a smooth open line γ in $\mathcal{M}$: $x = x(\xi) \to X(x(\xi))$, $(x(0) = x_0)$, by requiring $\mathbf{v}_{\|}(x_0) = \mathbf{v}(x_0)$ and[3]

$$\mathbf{v}_{\|}(x, \delta x) = \mathbf{v}_{\|}(x + \delta x) \tag{5.21}$$

for all $x \in \gamma$ and $\delta x = \delta\xi(dx/d\xi)$ a small shift along γ.

According to equation (5.19), the condition (5.21) of autoparallelity for all $x \in \gamma$ requires

$$\begin{aligned} v_{\|}^{\nu}(x) - \Gamma^{\nu}{}_{\rho\mu}(x)v_{\|}^{\rho}(x)\delta\xi\frac{dx^{\mu}}{d\xi} &= v_{\|}^{\nu}(x + \delta\xi(dx/d\xi)) \\ &= v_{\|}^{\nu}(x) + \delta\xi\frac{dx^{\mu}}{d\xi}\partial_{\mu}v_{\|}^{\nu}(x), \end{aligned} \tag{5.22}$$

i.e.

$$D_{\xi}v_{\|}^{\nu}(x) = \frac{dx^{\mu}}{d\xi}D_{\mu}v_{\|}^{\nu}(x) = \frac{dx^{\mu}}{d\xi}[\partial_{\mu}v_{\|}^{\nu}(x) + \Gamma^{\nu}{}_{\rho\mu}(x)v_{\|}^{\rho}(x)] = 0. \tag{5.23}$$

We can integrate this equation[4] along γ:

$$\begin{aligned} v_{\|}^{\nu}(x_1) = v^{\nu}(x_0) &+ \int_0^{\xi_1} d\xi\frac{dx^{\mu}}{d\xi}\partial_{\mu}v_{\|}^{\nu}(x)\bigg|_{x=x(\xi)} \\ &= v^{\nu}(x_0) - \int_0^{\xi_1} d\xi\frac{dx^{\mu}}{d\xi}\Gamma^{\nu}{}_{\rho\mu}(x)v_{\|}^{\rho}(x)\bigg|_{x=x(\xi)}. \end{aligned} \tag{5.25}$$

If the curve γ is only piecewise smooth (implying continuity in our notion of piecewise smoothness), then $\mathbf{v}(x_0)$ is parallel translated to become $\mathbf{v}_{\|}(x)$ if equation (5.23) holds along any smooth piece of the curve γ.

Since the parallelity condition (5.21) can also equivalently be expressed in terms of the covariant vector components,

$$D_{\xi}v_{\|\nu}(x) = \frac{dx^{\mu}}{d\xi}D_{\mu}v_{\|\nu}(x) = \frac{dx^{\mu}}{d\xi}[\partial_{\mu}v_{\|\nu}(x) - v_{\|\rho}(x)\Gamma^{\rho}{}_{\nu\mu}(x)] = 0, \tag{5.26}$$

we find that parallel transport of vector fields preserves scalar products:

---

[3] Levi-Cività T 1917 *Rendiconti del Circolo Matematico di Palermo* 42 173.

[4] If you are familiar with advanced quantum mechanics, you may recognize that iteration of the right-hand side of equation (5.25) yields an expression similar to the time-evolution operators in quantum mechanics,

$$v_{\|}(x_1) = \mathrm{P}\exp\left(-\int_{\gamma} dx^{\mu}\,\underline{\Gamma}_{\mu}(x)\right)\cdot v(x_0), \tag{5.24}$$

where the transport operator on the right-hand side is a path-ordered exponential. See e.g. chapter 13 in Dick R 2016 *Advanced Quantum Mechanics: Materials and Photons* 2nd edn (Cham: Springer). However, we do not have to use the integration (5.24) of equation (5.25) in this course.

$$\frac{d}{d\xi}(w_{\|\nu}v_{\|}^{\nu}) = \frac{dx^{\mu}}{d\xi}\partial_{\mu}(w_{\|\nu}v_{\|}^{\nu}) = w_{\|\rho}\Gamma^{\rho}{}_{\nu\mu}v_{\|}^{\nu} - w_{\|\nu}\Gamma^{\nu}{}_{\rho\mu}v_{\|}^{\rho} = 0. \tag{5.27}$$

5.3 Transformations of tensors and Christoffel symbols

The chain rule for differentiation implies under coordinate transformations $x^{\alpha} \to x'^{\mu}(x)$ the following transformation property of basis vectors,

$$\boldsymbol{E}_{\alpha}(x) = \partial_{\alpha}\boldsymbol{r} \to \boldsymbol{E}'_{\mu}(x') = \partial'_{\mu}\boldsymbol{r} = \partial'_{\mu}x^{\alpha}\,\partial_{\alpha}\boldsymbol{r} = \partial'_{\mu}x^{\alpha}\,\boldsymbol{E}_{\alpha}(x). \tag{5.28}$$

This transformation behavior of the basis vectors is denoted as *covariant*. It is inverse to the contravariant transformation behavior of coordinate shifts:

$$\delta x^{\alpha} \to \delta x'^{\mu} = \delta x^{\alpha}\partial_{\alpha}x'^{\mu}. \tag{5.29}$$

We encountered covariant and contravariant transformations already in chapter 2. The more general definitions here reduce to the definitions in chapter 2 if we specialize to linear transformations

$$x^{\alpha} = M^{\alpha}{}_{\mu}x'^{\mu} + C^{\alpha}, \quad x'^{\mu} = (M^{-1})^{\mu}{}_{\alpha}(x^{\alpha} - C^{\alpha}). \tag{5.30}$$

Invariance of vectors and of the requirements for dual vectors under coordinate transformations,

$$v'^{\mu}\boldsymbol{E}'_{\mu} = v^{\alpha}\boldsymbol{E}_{\alpha}, \quad \boldsymbol{E}'^{\mu}\cdot\boldsymbol{E}'_{\nu} = \delta^{\mu}{}_{\nu} = \boldsymbol{E}^{\mu}\cdot\boldsymbol{E}_{\nu} \tag{5.31}$$

imply that the expansion coefficients of vectors v^{μ} and the dual vectors $\boldsymbol{E}^{\mu}$ transform contravariantly:

$$v'^{\mu}(x') = v^{\alpha}(x)\partial_{\alpha}x'^{\mu}, \quad \boldsymbol{E}'^{\mu}(x') = \boldsymbol{E}^{\alpha}(x)\partial_{\alpha}x'^{\mu}. \tag{5.32}$$

As a rule upper indices transform with the Jacobian matrix $\partial_{\alpha}x'^{\mu}$, and lower indices with the inverse Jacobian matrix $\partial'_{\mu}x^{\alpha}$ of the coordinate transformation, e.g.

$$T'_{\mu}{}^{\nu} = \partial'_{\mu}x^{\alpha}\,T_{\alpha}{}^{\beta}\,\partial_{\beta}x'^{\nu}. \tag{5.33}$$

Objects with such a transformation behavior under coordinate transformations are denoted as *tensors*. Note that a tensor vanishes in one coordinate system if and only if it vanishes in every coordinate system.

We have seen that vectors are tensors, and the metric is also a tensor:

$$g'_{\mu\nu} = \boldsymbol{E}'_{\mu}\cdot\boldsymbol{E}'_{\nu} = \partial'_{\mu}x^{\alpha}\,\partial'_{\nu}x^{\beta}\,g_{\alpha\beta}, \tag{5.34}$$

but Christoffel symbols are not tensors:

Transformation of Christoffel symbols

From $\boldsymbol{E}'_{\mu}(x') = \partial'_{\mu}x^{\alpha}\,\boldsymbol{E}_{\alpha}(x)$, and $\boldsymbol{E}'^{\mu}(x') = \boldsymbol{E}^{\alpha}(x)\partial_{\alpha}x'^{\mu}$ follows the transformation law for Christoffel symbols,

$$\Gamma'^{\nu}{}_{\rho\mu} = \boldsymbol{E}'^{\nu} \cdot \partial'_{\mu}\boldsymbol{E}'_{\rho} = \boldsymbol{E}^{\alpha}\partial_{\alpha}x'^{\nu} \cdot \partial'_{\mu}x^{\beta}\partial_{\beta}(\partial'_{\rho}x^{\gamma}\,\boldsymbol{E}_{\gamma})$$
$$= \partial_{\alpha}x'^{\nu}\,\Gamma^{\alpha}{}_{\gamma\beta}\partial'_{\rho}x^{\gamma}\,\partial'_{\mu}x^{\beta} + \partial_{\alpha}x'^{\nu}\,\partial'_{\mu}\partial'_{\rho}x^{\alpha}. \tag{5.35}$$

This inhomogeneous transformation behavior implies that even if the Christoffel symbols vanish in one coordinate system (this is the case for straight coordinate lines in a flat space), then just transforming the coordinates nonlinearly (such that $\partial'_{\mu}\partial'_{\rho}x^{\alpha} \neq 0$) will yield non-vanishing Christoffel symbols. Christoffel symbols therefore also appear in flat spaces if one uses curvilinear coordinates, see e.g. equation (2.81) for the non-vanishing Christoffel symbols for polar coordinates in the plane.

The partial derivatives of the components of a vector also do not correspond to a tensor,

$$\partial'_{\mu}v'^{\nu} = \partial'_{\mu}x^{\alpha}\,\partial_{\alpha}(v^{\beta}\,\partial_{\beta}x'^{\nu}) = \partial'_{\mu}x^{\alpha}\,\partial_{\alpha}v^{\beta}\,\partial_{\beta}x'^{\nu} + \partial'_{\mu}x^{\alpha}\,v^{\beta}\,\partial_{\alpha}\partial_{\beta}x'^{\nu}. \tag{5.36}$$

However, the extra terms in equations (5.35) and (5.36) cancel in such a way that the covariant derivatives $D_{\mu}v^{\nu}$ transform like the components of a tensor,

$$D'_{\mu}v'^{\nu} = \partial'_{\mu}v'^{\nu} + \Gamma'^{\nu}{}_{\rho\mu}v'^{\rho} = \partial'_{\mu}x^{\alpha}\,D_{\alpha}v^{\beta}\,\partial_{\beta}x'^{\nu}, \tag{5.37}$$

corresponding to the invariance of $D_{\delta x}\boldsymbol{v} = \delta x^{\alpha}D_{\alpha}v^{\beta}\boldsymbol{E}_{\beta}$. This property is the reason for the designation *covariant derivative*.

The verification of the covariance property (5.37) goes as follows:

$$\begin{aligned}
D'_{\mu}v'^{\nu} &= \partial'_{\mu}v'^{\nu} + \Gamma'^{\nu}{}_{\rho\mu}v'^{\rho} \\
&= \partial'_{\mu}x^{\alpha}\,\partial_{\alpha}(v^{\beta}\,\partial_{\beta}x'^{\nu}) + (\partial_{\alpha}x'^{\nu}\,\Gamma^{\alpha}{}_{\gamma\beta}\partial'_{\rho}x^{\gamma}\,\partial'_{\mu}x^{\beta} + \partial_{\alpha}x'^{\nu}\,\partial'_{\mu}\partial'_{\rho}x^{\alpha})v'^{\rho} \\
&= \partial'_{\mu}x^{\alpha}\,\partial_{\alpha}v^{\beta}\,\partial_{\beta}x'^{\nu} + \partial'_{\mu}x^{\alpha}\,v^{\beta}\,\partial_{\alpha}\partial_{\beta}x'^{\nu} + \partial_{\alpha}x'^{\nu}\,\Gamma^{\alpha}{}_{\gamma\beta}v^{\gamma}\partial'_{\mu}x^{\beta} \\
&\quad - \partial'_{\mu}\partial_{\alpha}x'^{\nu}\,v'^{\rho}\partial'_{\rho}x^{\alpha} \\
&= \partial'_{\mu}x^{\alpha}(\partial_{\alpha}v^{\beta} + \Gamma^{\beta}{}_{\gamma\alpha}v^{\gamma})\partial_{\beta}x'^{\nu} + v^{\beta}(\partial'_{\mu}x^{\alpha}\,\partial_{\alpha}\partial_{\beta}x'^{\nu} - \partial'_{\mu}\partial_{\beta}x'^{\nu}) \\
&= \partial'_{\mu}x^{\alpha}(\partial_{\alpha}v^{\beta} + \Gamma^{\beta}{}_{\gamma\alpha}v^{\gamma})\partial_{\beta}x'^{\nu} = \partial'_{\mu}x^{\alpha}\,D_{\alpha}v^{\beta}\,\partial_{\beta}x'^{\nu}.
\end{aligned} \tag{5.38}$$

In the previous equation the identities

$$\partial'_{\mu}(\partial'_{\rho}x^{\alpha}\,\partial_{\alpha}x'^{\nu}) = \partial'_{\mu}\delta_{\rho}{}^{\nu} = 0 \quad \Rightarrow \quad \partial_{\alpha}x'^{\nu}\,\partial'_{\mu}\partial'_{\rho}x^{\alpha} = -\partial'_{\mu}\partial_{\alpha}x'^{\nu}\,\partial'_{\rho}x^{\alpha} \tag{5.39}$$

and

$$\partial'_{\mu}x^{\alpha}\,\partial_{\alpha} = \partial'_{\mu} \quad \Rightarrow \quad \partial'_{\mu}x^{\alpha}\,\partial_{\alpha}\partial_{\beta}x'^{\nu} = \partial'_{\mu}\partial_{\beta}x'^{\nu} \tag{5.40}$$

were used.

The covariant derivatives of higher-order tensors can be read off from

$$D_{\mu}T_{\nu}{}^{\lambda} = \partial_{\mu}T_{\nu}{}^{\lambda} - \Gamma^{\rho}{}_{\nu\mu}T_{\rho}{}^{\lambda} + \Gamma^{\lambda}{}_{\rho\mu}T_{\nu}{}^{\rho}. \tag{5.41}$$

Each upper index yields a term which is contracted with $+\Gamma$, and each lower index yields a term which is contracted with $-\Gamma$.

A particular case is the covariant derivative of the metric. Substitution of the last equation in (5.13) shows that this vanishes identically,

$$D_\mu g_{\rho\sigma} = \partial_\mu g_{\rho\sigma} - \Gamma^\nu{}_{\rho\mu} g_{\nu\sigma} - \Gamma^\nu{}_{\sigma\mu} g_{\rho\nu} = 0, \tag{5.42}$$

$$D_\mu g^{\rho\sigma} = \partial_\mu g^{\rho\sigma} + \Gamma^\rho{}_{\nu\mu} g^{\nu\sigma} + \Gamma^\sigma{}_{\nu\mu} g^{\rho\nu} = 0. \tag{5.43}$$

Equation (5.41) is also consistent with a product rule for covariant derivatives, which is expected to hold due to its derivation (5.12) from first-order shifts[5]. You can easily verify that the equations (5.15), (5.16) and (5.41) comply with

$$D_\mu(v_\nu w^\lambda) = (D_\mu v_\nu) w^\lambda + v_\nu D_\mu w^\lambda. \tag{5.44}$$

[5] At a formal level, the result (5.41) for a general tensor follows from direct generalization of the definitions (5.11) and (5.12) to general $(m+n)$th-order tensors $\underline{T} = T^{\alpha_1 \ldots \alpha_m}{}_{\beta_1 \ldots \beta_n} \boldsymbol{E}_{\alpha_1} \otimes \ldots \boldsymbol{E}_{\alpha_m} \otimes \boldsymbol{E}^{\beta_1} \otimes \ldots \boldsymbol{E}^{\beta_n}$, and comparison with the covariant derivatives of tensor products of lower-order tensors then proves the general product rule for covariant derivatives.

Chapter 6

Particles in curved spacetime

We are now prepared to generalize the relativistic equations of motion, which we have discovered in inertial frames in the Minkowski spacetime, to the case of curved spacetimes. Indeed, the equations of motion for electromagnetic fields and for particles look the same in curvilinear coordinates in Minkowski spacetime and in curved spacetimes, and we cannot tell from those equations alone whether we are dealing with motion in flat Minkowski spacetime or in curved spacetimes. This is a manifestation of the *equivalence of inertial forces and gravitational forces for the motion of individual particles*. The physical distinction between curvilinear coordinates in Minkowski spacetime on the one hand and curved spacetime on the other hand involves the motion of ensembles of particles and will be discussed in chapter 7.

6.1 Motion of a particle in spacetime

We have seen in equations (4.78) and (4.80) that the action of a free particle in flat Minkowski spacetime and in inertial coordinates is up to a factor $-mc$ just the length of the world line of the particle,

$$\begin{aligned}S &= -mc\int dt\sqrt{c^2-\dot{\boldsymbol{x}}^2(t)} = -mc\int\sqrt{-\eta_{\mu\nu}dx^\mu dx^\nu}\\ &= -mc\int d\xi\sqrt{-\eta_{\mu\nu}(dx^\mu/d\xi)(dx^\nu/d\xi)}\,.\end{aligned}$$

In the last equation we wrote this with an arbitrary parameter ξ along the world line of the particle, because nothing requires us to use our lab time or any other particular time to label the points $x(\xi)$ along the world line.

On the other hand, nothing (besides convenience) forces us to use inertial coordinates to label points in the Minkowski spacetime. We could just as well use curvilinear coordinates. If the coordinates y^μ are curvilinear coordinates, then the length element along the world line is given by

doi:10.1088/2053-2571/aaf173ch6

$$
\begin{aligned}
c^2 d\tau^2 &= -\eta_{\alpha\beta} dx^\alpha dx^\beta = -\eta_{\alpha\beta} \frac{\partial x^\alpha(y)}{\partial y^\mu} \frac{\partial x^\beta(y)}{\partial y^\nu} dy^\mu dy^\nu \\
&= -g_{\mu\nu}(y) dy^\mu dy^\nu,
\end{aligned}
\tag{6.1}
$$

and the action of the free particle is

$$
S = -mc \int d\xi \sqrt{-g_{\mu\nu}(y(\xi))(dy^\mu(\xi)/d\xi)(dy^\nu(\xi)/d\xi)}\,. \tag{6.2}
$$

Nothing in this expression reminds us of flat Minkowski spacetime any more, and the same expression would hold in this form also in a curved spacetime. Therefore this expression should also describe the action of a free particle in a curved spacetime.

Following standard conventions, we will denote the coordinates again with x^μ in the following. The action of a free particle in flat or curved spacetime is then

$$
S = -mc \int d\xi \sqrt{-g_{\mu\nu}(x(\xi))\dot{x}^\mu(\xi)\dot{x}^\nu(\xi)}\,, \tag{6.3}
$$

where we used the definition $\dot{x}^\mu(\xi) \equiv dx^\mu(\xi)/d\xi$.

The meaning of the metric and clocks in spacetime

So far we have just used the geometric notion of the metric

$$
ds^2 = g_{\mu\nu}(x) dx^\mu dx^\nu \tag{6.4}
$$

that it gives us distances between points in a space. In a spacetime, $\sqrt{-ds^2}$ is the distance between the two nearby points with coordinates x^μ and $x^\mu + dx^\mu$, where we assume timelike separation of the two points, $ds^2 < 0$. This is directly connected to the physics of general relativity through the definition of ideal clocks: An *ideal clock* is a clock which measures a time

$$
d\tau = \frac{1}{c}\sqrt{-ds^2} = \frac{1}{c}\sqrt{-g_{\mu\nu}(x) dx^\mu dx^\nu} \tag{6.5}
$$

when it moves from the point with coordinates x^μ to the point with coordinates $x^\mu + dx^\mu$. For any object moving along a timelike path $x^\mu = x^\mu(\xi)$, the time

$$
\tau = \frac{1}{c}\int_{\xi_1}^{\xi_2} d\xi \sqrt{-g_{\mu\nu}(x(\xi))\dot{x}^\mu(\xi)\dot{x}^\nu(\xi)} \tag{6.6}
$$

is the *eigentime* of that object while it moves from $x(\xi_1)$ to $x(\xi_2)$, and it is the time which would have elapsed on an ideal clock traveling with the object.

Equations of motion for free particles

Application of the principle of stationary action, $\delta S = 0$, to equation (6.3) yields the equations of motion

$$\left(\frac{\partial}{\partial x^\mu} - \frac{d}{d\xi}\frac{\partial}{\partial \dot{x}^\mu}\right)\sqrt{-g_{\alpha\beta}(x)\dot{x}^\alpha \dot{x}^\beta} = \frac{d}{d\xi}\left(\frac{1}{\sqrt{-g_{\alpha\beta}\dot{x}^\alpha\dot{x}^\beta}} g_{\mu\nu}\dot{x}^\nu\right) - \frac{1}{2\sqrt{-g_{\alpha\beta}\dot{x}^\alpha\dot{x}^\beta}}\partial_\mu g_{\rho\sigma}\dot{x}^\rho\dot{x}^\sigma \tag{6.7}$$

$$= \frac{1}{\sqrt{-g_{\alpha\beta}\dot{x}^\alpha\dot{x}^\beta}}\left[g_{\mu\nu}\ddot{x}^\nu + \partial_\rho g_{\mu\nu}\dot{x}^\rho\dot{x}^\nu - \frac{1}{g_{\gamma\delta}\dot{x}^\gamma\dot{x}^\delta}\left(g_{\rho\sigma}\dot{x}^\rho\ddot{x}^\sigma + \frac{1}{2}\partial_\lambda g_{\rho\sigma}\dot{x}^\lambda\dot{x}^\rho\dot{x}^\sigma\right)g_{\mu\nu}\dot{x}^\nu - \frac{1}{2}\partial_\mu\, g_{\rho\sigma}\dot{x}^\rho\dot{x}^\sigma\right] = 0, \tag{6.8}$$

i.e.

$$\ddot{x}^\mu + g^{\mu\sigma}\left(\partial_\rho g_{\sigma\nu} - \frac{1}{2}\partial_\sigma g_{\rho\nu}\right)\dot{x}^\rho\dot{x}^\nu = \frac{1}{g_{\gamma\delta}\dot{x}^\gamma\dot{x}^\delta}\left(g_{\rho\sigma}\dot{x}^\rho\ddot{x}^\sigma + \frac{1}{2}\partial_\lambda g_{\rho\sigma}\dot{x}^\lambda\dot{x}^\rho\dot{x}^\sigma\right)\dot{x}^\mu. \tag{6.9}$$

This yields with

$$\Gamma^\mu{}_{\nu\rho}\dot{x}^\nu\dot{x}^\rho = g^{\mu\sigma}\left(\partial_\rho g_{\sigma\nu} - \frac{1}{2}\partial_\sigma g_{\rho\nu}\right)\dot{x}^\rho\dot{x}^\nu \tag{6.10}$$

the following equation

$$\ddot{x}^\mu + \Gamma^\mu{}_{\nu\rho}\dot{x}^\nu\dot{x}^\rho = \frac{1}{\dot{x}^2}\dot{x}^\mu\dot{x}^\nu g_{\nu\rho}(\ddot{x}^\rho + \Gamma^\rho{}_{\lambda\sigma}\dot{x}^\lambda\dot{x}^\sigma), \tag{6.11}$$

which tells us that $\ddot{x}^\mu + \Gamma^\mu{}_{\nu\rho}\dot{x}^\nu\dot{x}^\rho \propto \dot{x}^\mu$. However, we find a more useful form by writing equation (6.8) in the following form:

$$\left(\frac{\partial}{\partial x^\mu} - \frac{d}{d\xi}\frac{\partial}{\partial \dot{x}^\mu}\right)\sqrt{-g_{\alpha\beta}\dot{x}^\alpha \dot{x}^\beta} = \frac{d}{d\xi}\left(\frac{1}{\sqrt{-g_{\alpha\beta}\dot{x}^\alpha\dot{x}^\beta}} g_{\mu\nu}\dot{x}^\nu\right) - \frac{1}{2\sqrt{-g_{\alpha\beta}\dot{x}^\alpha\dot{x}^\beta}}\partial_\mu g_{\rho\sigma}\dot{x}^\rho\dot{x}^\sigma$$
$$= \frac{1}{\sqrt{-g_{\alpha\beta}\dot{x}^\alpha\dot{x}^\beta}}\left[g_{\mu\nu}\ddot{x}^\nu + \partial_\rho g_{\mu\nu}\dot{x}^\rho\dot{x}^\nu - g_{\mu\nu}\dot{x}^\nu\frac{d}{d\xi}\ln\left(\frac{1}{c}\sqrt{-g_{\gamma\delta}\dot{x}^\gamma\dot{x}^\delta}\right) - \frac{1}{2}\partial_\mu g_{\rho\sigma}\dot{x}^\rho\dot{x}^\sigma\right] = 0. \tag{6.12}$$

This yields[1]

[1] The constant factor $1/c$ cancels in the calculation of the derivative and is only introduced to make the argument of the logarithm dimensionless.

$$\ddot{x}^\mu + \Gamma^\mu{}_{\nu\rho}\dot{x}^\nu\dot{x}^\rho = \dot{x}^\mu \frac{d}{d\xi}\ln\left(\frac{1}{c}\sqrt{-g_{\gamma\delta}\dot{x}^\gamma\dot{x}^\delta}\right). \tag{6.13}$$

Comparison with equation (5.17) shows that the left-hand side of equation (6.13) is just the covariant derivative of the 4-velocity of the particle along the world line of the particle: $D_\xi\dot{x}^\mu = \ddot{x}^\mu + \Gamma^\mu{}_{\rho\nu}\dot{x}^\rho\dot{x}^\nu$.

We can simplify equation (6.13) by noting that

$$cd\tau = \sqrt{-g_{\mu\nu}(x)dx^\mu dx^\nu} = \sqrt{-g_{\mu\nu}(x)\frac{dx^\mu}{d\xi}\frac{dx^\nu}{d\xi}}d\xi. \tag{6.14}$$

Therefore, if we choose the particle's eigentime to label the points along its world line, $\xi = \tau$, we have

$$\sqrt{-g_{\mu\nu}(x(\tau))\dot{x}^\mu(\tau)\dot{x}^\nu(\tau)} = c, \tag{6.15}$$

and equation (6.13) simplifies to

$$D_\tau\dot{x}^\mu = \ddot{x}^\mu + \Gamma^\mu{}_{\nu\rho}\dot{x}^\nu\dot{x}^\rho = 0. \tag{6.16}$$

Sometimes we prefer to have the equations of motion written with respect to the coordinate time t. We can multiply equation (6.16) with

$$\frac{d\tau}{dt} = \frac{1}{c}\sqrt{-g_{\mu\nu}\frac{dx^\mu}{dt}\frac{dx^\nu}{dt}} = \frac{1}{c}\sqrt{-g_{00}c^2 - 2g_{0i}v^i - g_{ij}v^i v^j} \tag{6.17}$$

to convert one τ-derivative into a time derivative, and use the definition of the 4-momentum

$$p^\mu = m\frac{dx^\mu}{d\tau} = \frac{mc}{\sqrt{-g_{00}c^2 - 2g_{0i}v^i - g_{ij}v^i v^j}}\frac{dx^\mu}{dt}. \tag{6.18}$$

The equation in terms of the coordinate time t is then

$$\frac{d}{dt}p^\mu + \Gamma^\mu{}_{\nu\rho}p^\nu\frac{x^\rho}{dt} = 0. \tag{6.19}$$

Equations (6.16) or (6.19) are the equations of motion of a particle in the absence of electromagnetic (and weak and strong) forces acting on the particle, if we use eigentime τ or the coordinate time t to label points along the world line of the particle. The left-hand side of equation (6.16) is the covariant derivative of the particle's 4-velocity along its world line, and is the *covariant acceleration* of the particle.

Note that the Christoffel symbols in equations (6.16) or (6.19) only came from the x-dependence $g_{\mu\nu}(x)$ of the metric. This x-dependence can have two reasons: It can arise due to the use of curvilinear coordinates in a flat spacetime, or it can be unavoidable due to the curvature of spacetime. On the level of the single particle equation of motion this is not distinguishable: The Christoffel symbols in equation

(6.19) can correspond to inertial forces due to the use of curvilinear coordinates or gravitational forces due to the curvature of spacetime[2].

Equation (6.16) is the necessary and sufficient condition for stationary length of a smooth path with fixed endpoints, if the eigentime (or equivalently the length of the path) is used as a parameter. Therefore, it is also a necessary (but not sufficient) condition for the smooth path to have extremal length. For this reason equation (6.16) is denoted as the *geodesic equation*.

Equation (6.16) is the standard statement of the geodesic equation. However, note that this equation also implies

$$\begin{aligned}\frac{d}{d\tau}\left(g_{\mu\nu}\dot{x}^\nu\right) &= \dot{x}^\nu\dot{x}^\sigma\partial_\sigma g_{\mu\nu} + g_{\mu\nu}\ddot{x}^\nu = \dot{x}^\nu\dot{x}^\sigma\partial_\sigma g_{\mu\nu} - g_{\mu\nu}\Gamma^\nu{}_{\rho\sigma}\dot{x}^\rho\dot{x}^\sigma \\ &= \frac{1}{2}\dot{x}^\rho\dot{x}^\sigma\partial_\mu g_{\rho\sigma},\end{aligned} \tag{6.20}$$

i.e. for the covariant components of the 4-momentum $p_\mu = mg_{\mu\nu}\dot{x}^\nu$

$$\frac{d}{d\tau}p_\mu(\tau) = \frac{m}{2}\dot{x}^\rho(\tau)\dot{x}^\sigma(\tau)\partial_\mu g_{\rho\sigma}(x(\tau)). \tag{6.21}$$

If we use the particle's eigentime τ to parametrize its world line, we have from $L = -mc\sqrt{-\dot{x}^2(\tau)} = -mc^2$ and with the definition[3] $\dot{x}_\mu \equiv g_{\mu\nu}(x)\dot{x}^\nu$,

$$p_\mu = \frac{\partial L}{\partial \dot{x}^\mu} = mc\frac{\dot{x}_\mu(\tau)}{\sqrt{-\dot{x}^2(\tau)}} = m\dot{x}_\mu(\tau). \tag{6.22}$$

The momentum and the equation (6.16) in eigentime gauge also follow from the equivalent Lagrange function

$$L_\tau = \frac{m}{2}\dot{x}^2(\tau) = -\frac{m}{2}c^2. \tag{6.23}$$

For the proof that L_τ yields the correct equation of motion we note

$$\begin{aligned}&\frac{1}{2}g^{\mu\nu}(x)\left(\frac{d}{d\tau}\frac{\partial}{\partial\dot{x}^\nu} - \frac{\partial}{\partial x^\nu}\right)g_{\kappa\lambda}(x)\dot{x}^\kappa\dot{x}^\lambda = g^{\mu\nu}(x)\frac{d}{d\tau}(g_{\nu\lambda}(x)\dot{x}^\lambda) \\ &- \frac{1}{2}g^{\mu\nu}(x)\partial_\nu g_{\kappa\lambda}(x)\dot{x}^\kappa\dot{x}^\lambda = \ddot{x}^\mu + \dot{x}^\kappa\dot{x}^\lambda g^{\mu\nu}(x)\left(\partial_\kappa g_{\nu\lambda}(x) - \frac{1}{2}\partial_\nu g_{\kappa\lambda}(x)\right) \\ &= \ddot{x}^\mu + \Gamma^\mu{}_{\kappa\lambda}(x)\dot{x}^\kappa\dot{x}^\lambda.\end{aligned} \tag{6.24}$$

If we use eigentime τ, this is the most direct way to derive the left-hand side of the equations of motion (6.16) of a particle in a space with metric $g_{\mu\nu}(x)$. This also

[2] …in a *conventional mechanical interpretation*. However, from a puristic or geometric point of view, the term with the Christoffel symbols should *not* be denoted as a force at all: It is just the necessary modification of the definition of acceleration due to the presence of curvature, or due to the use of curvilinear coordinates, respectively.

[3] Note that with the given definition, $\dot{x}_\mu$ is generically *not* the same as $dx_\mu(\tau)/d\tau$.

implies that the derivation of the geodesic equation from the eigentime Lagrange function (6.23) is the simplest and most direct way to calculate the Christoffel symbols of a metric analytically.

Equation of motion with electromagnetic forces

Here we use $\xi = \tau$ right away to abbreviate the calculation. The action of a charged particle in the presence of electromagnetic potentials is

$$\begin{aligned} S &= -mc\int d\tau\sqrt{-g_{\mu\nu}(x)\dot{x}^\mu\dot{x}^\nu} + q\int d\tau\ \dot{x}^\mu A_\mu(x) \\ &= -mc^2\int d\tau + q\int d\tau\ \dot{x}^\mu A_\mu(x), \end{aligned} \tag{6.25}$$

and the corresponding Euler–Lagrange equations are

$$\begin{aligned} &\left(\frac{\partial}{\partial x^\mu} - \frac{d}{d\tau}\frac{\partial}{\partial\dot{x}^\mu}\right)\left(\sqrt{-g_{\alpha\beta}(x)\dot{x}^\alpha\dot{x}^\beta} - \frac{q}{mc}\dot{x}^\alpha A_\alpha(x)\right) \\ &= \frac{d}{d\tau}\left(\frac{1}{c}g_{\mu\nu}\dot{x}^\nu + \frac{q}{mc}A_\mu\right) - \frac{1}{2c}\partial_\mu g_{\rho\sigma}\dot{x}^\rho\dot{x}^\sigma - \frac{q}{mc}\dot{x}^\alpha\partial_\mu A_\alpha(x) \\ &= \frac{1}{c}\left(g_{\mu\nu}\ddot{x}^\nu + \partial_\rho g_{\mu\nu}\dot{x}^\rho\dot{x}^\nu + \frac{q}{m}\dot{x}^\nu\partial_\nu A_\mu - \frac{1}{2}\partial_\mu g_{\rho\sigma}\dot{x}^\rho\dot{x}^\sigma - \frac{q}{m}\dot{x}^\nu\partial_\mu A_\nu(x)\right) \\ &= 0. \end{aligned} \tag{6.26}$$

The terms containing derivatives of the metric yield a Christoffel symbol, and the terms with derivatives of the electromagnetic potentials yield the field strength tensor, i.e. the equations of motion are

$$\ddot{x}^\mu + \Gamma^\mu{}_{\rho\nu}\dot{x}^\rho\dot{x}^\nu = \frac{q}{m}F^\mu{}_\nu\dot{x}^\nu. \tag{6.27}$$

The equations in terms of the coordinate time t are then

$$\frac{d}{dt}p^\mu + \Gamma^\mu{}_{\rho\nu}p^\rho\frac{dx^\nu}{dt} = qF^\mu{}_\nu\frac{dx^\nu}{dt}. \tag{6.28}$$

Just like for the free particle (6.23), we would receive the same equations of motion (6.27) from the simpler Lagrange function

$$L_\tau = \frac{m}{2}\dot{x}^2(\tau) + q\dot{x}(\tau)\cdot A(x(\tau)), \tag{6.29}$$

but recall that this requires the use of eigentime to parametrize the particle's world line.

6.2 Slow particles in a weak gravitational field

Now we will examine what the equation for the world line of a free particle

$$D_\tau x^\mu = \ddot{x}^\mu + \Gamma^\mu{}_{\rho\nu}\dot{x}^\rho\dot{x}^\nu = 0 \tag{6.30}$$

implies for slow particles in an approximately flat spacetime.

Approximate flatness means that we can choose coordinates such that the metric approximates the flat Minkowski metric along the whole piece of the particle trajectory which we are considering:

$$g_{\mu\nu}(x) = \eta_{\mu\nu} + h_{\mu\nu}(x) \tag{6.31}$$

with

$$|h_{\mu\nu}(x)| \ll 1. \tag{6.32}$$

Slow means

$$\left| \frac{dx^i}{dt} \right| \ll c. \tag{6.33}$$

The equations (6.31)–(6.33) imply for the eigentime of the moving particle in first order

$$\begin{aligned} cd\tau = dt\sqrt{-g_{00}c^2 - 2g_{0i}v^i - g_{ij}v^i v^j} &\simeq cdt\sqrt{1 - h_{00}} \\ \simeq \left(1 - \frac{1}{2}h_{00}\right) cdt, \end{aligned} \tag{6.34}$$

and

$$\frac{d}{d\tau} \simeq \left(1 + \frac{1}{2}h_{00}\right) \frac{d}{dt}. \tag{6.35}$$

In first order the spatial velocity components of the moving particle have disappeared from the relation between τ and t. These two time parameters have the following meaning: τ is the time measured by a clock moving with the particle in the background metric $g_{\mu\nu}(x) = \eta_{\mu\nu} + h_{\mu\nu}(x)$, while t is the time measured by a clock which is at rest in the coordinate system in a point x where $h_{00}(x) = 0$. In chapter 7 we will find that this typically requires the clock measuring the time t to sit at spatial infinity.

The first-order expansion of the Christoffel symbols yields

$$\Gamma^{\mu}{}_{\rho\nu} \simeq \frac{1}{2}\eta^{\mu\sigma} \left(\partial_\rho h_{\sigma\nu} + \partial_\nu h_{\sigma\rho} - \partial_\sigma h_{\rho\nu}\right), \tag{6.36}$$

and therefore we have in first order

$$\Gamma^{\mu}{}_{\rho\nu}\dot{x}^\rho \dot{x}^\nu \simeq \Gamma^{\mu}{}_{00}c^2 \simeq \eta^{\mu\sigma}\left(\partial_0 h_{\sigma 0} - \frac{1}{2}\partial_\sigma h_{00}\right)c^2. \tag{6.37}$$

The geodesic equation for $\mu = i \in \{1, 2, 3\}$ then becomes in first order

$$\frac{d^2x^i}{dt^2} + \frac{\partial}{\partial t}h_{i0}c - \frac{1}{2}\partial_i h_{00}c^2 = 0, \tag{6.38}$$

while the equation for $\mu = 0$ becomes trivial in lowest order,

$$\frac{d}{dt}\left(1 + \frac{1}{2}h_{00}\right)c + \Gamma^0{}_{00}c^2 = \frac{1}{2}\frac{d}{dt}h_{00}c - \frac{1}{2}\frac{\partial}{\partial t}h_{00}c = 0, \tag{6.39}$$

since

$$\frac{d}{dt}h_{00} - \frac{\partial}{\partial t}h_{00} = v^i\partial_i h_{00} \tag{6.40}$$

is already a second-order term.

Equation (6.38) tells us that in a constant or approximately time-independent field, $\partial_t h_{i0} = 0$, the quantity $h_{00}(x) = g_{00}(x) - \eta_{00} = g_{00}(x) + 1$ corresponds to the gravitational potential Φ: Comparison of

$$\frac{d^2x^i}{dt^2} = \frac{1}{2}\partial_i h_{00}c^2 \tag{6.41}$$

with

$$\frac{d^2x^i}{dt^2} = -\partial_i\Phi \tag{6.42}$$

implies

$$h_{00}(\boldsymbol{x}) = -\frac{2}{c^2}\Phi(\boldsymbol{x}) = \frac{2G}{c^2}\int d^3\boldsymbol{x}'\frac{\varrho(\boldsymbol{x}')}{|\boldsymbol{x} - \boldsymbol{x}'|}. \tag{6.43}$$

The last equation assumes that the gravitational potential is generated by a mass density $\varrho(\boldsymbol{x})$.

The perturbation $g_{00} - \eta_{00}$ on Earth's surface is

$$h_{00} = \frac{2GM_\oplus}{c^2R_\oplus} = \frac{2 \times 6.673 \times 5.974 \times 10^{13}\,\mathrm{m}^3\,\mathrm{s}^{-2}}{2.998^2 \times 6.378 \times 10^{22}\,\mathrm{m}^3\,\mathrm{s}^{-2}} = 1.39 \times 10^{-9}. \tag{6.44}$$

The perturbation on the surface of the Sun is three orders of magnitude stronger, but still very small,

$$h_{00} = \frac{2GM_\odot}{c^2R_\odot} = \frac{2 \times 6.673 \times 1.989 \times 10^{19}}{2.998^2 \times 6.961 \times 10^{24}} = 4.24 \times 10^{-6}. \tag{6.45}$$

However, on the surface of a neutron star with $M = 1.4M_\odot$ and $R = 10^4$ m equation (6.43) would yield

$$h_{00} = \frac{2GM}{c^2R} = \frac{2 \times 6.673 \times 2.785 \times 10^{19}}{2.998^2 \times 10^{20}} = 0.414. \tag{6.46}$$

In this case the Newtonian approximation to gravity is apparently unsuitable, and we also cannot apply the relation (6.43) any more.

A few remarks on motion in weak gravitational fields:

1. The equation (6.38) is one part of the Newtonian limit of general relativity. The other part concerns the calculation of the metric $g_{\mu\nu}(x)$ in the weak field limit[4]. We can address this second part only after we have figured out the equations which determine the metric in general relativity in chapter 7.
2. We found

$$d\tau \simeq \left(1 - \frac{1}{2}h_{00}\right) dt = \left(1 + \frac{1}{c^2}\Phi\right) dt < dt, \tag{6.47}$$

i.e. clocks in a gravitational field are slowed down! The deeper the clock is in a gravitational potential well, the more it is slowed down. We will see later that this effect is not limited to weak gravitational fields.

3. The identification of h_{00} with the gravitational potential is much faster with the version (6.21) of the geodesic equation, if we use that in first order $d\boldsymbol{p}/d\tau = d\boldsymbol{p}/dt$,

$$\frac{d\boldsymbol{p}}{dt} = m\frac{c^2}{2}\nabla h_{00} = -m\nabla\Phi, \tag{6.48}$$

whence we find again $h_{00} = -2\Phi/c^2$.

6.3 Local inertial frames

Inertial frames in mechanics and physics before general relativity were understood as non-rotating coordinate systems in uniform motion relative to the 'rest frame of the fixed stars', and free particles were defined as particles which have constant velocity $\boldsymbol{v}$ in an inertial frame, thus satisfying the free Newton equation $d^2\boldsymbol{x}/dt^2 = \boldsymbol{0}$.

However, meanwhile we have learned that free particles are following curved paths if spacetime is curved,

$$\frac{d^2x^\mu}{d\tau^2} + \Gamma^\mu{}_{\rho\nu}\frac{dx^\rho}{d\tau}\frac{dx^\nu}{d\tau} = 0, \tag{6.49}$$

and this equation is generally covariant in that it holds in every coordinate system x^μ, without any reference to a cosmic rest frame or otherwise special coordinate systems.

How then do the inertial frames of pre-GTR physics fit into the picture? And how can we rediscover the free Newton equation for free particles, i.e. for freely falling particles? The following result holds.

If the curve $\tilde{\gamma}$ is a timelike geodesic through spacetime, then in a neighborhood $\mathcal{U}_{\tilde{\gamma}}$ of $\tilde{\gamma}$ it is possible to construct coordinates $\xi^a = (ct, \xi^{\bar{a}})$, $0 \leqslant a \leqslant 3$, $1 \leqslant \bar{a} \leqslant 3$, such

[4] Compare with electrodynamics. There are two parts to it: The equation of motion for charged particles in electromagnetic fields, and the Maxwell equations which describe the generation of electromagnetic fields from charges and currents. In this course, so far we only know the equation of motion for particles (6.16) or (6.27), but we have not yet talked about the equations which determine the gravitational fields (i.e. actually the metric of spacetime).

that up to terms of order $\mathcal{O}(|\boldsymbol{\xi}|)$ the four-dimensional geodesic equation in $\mathcal{U}_{\tilde{\gamma}}$ reduces to the three-dimensional free Newton equation $d^2\xi^{\bar{a}}/dt^2 = 0$, where t is the eigentime along the geodesic. This provides a *local* construction of inertial frames in the sense that it provides *local coordinates* in $\mathcal{U}_{\tilde{\gamma}}$ such that free particles satisfy the free Newton equation *in these coordinates*. However, these coordinate systems themselves generically are *not* in uniform motion relative to the cosmic rest frame.

For an example of a local inertial frame consider the trajectory of the center of mass of a space shuttle orbiting around Earth. The neighborhood $\mathcal{U}_{\tilde{\gamma}}$ traced out by the space shuttle itself is a good approximation of a local inertial frame, with $ct = \xi^0$ being the eigentime of the space shuttle as measured by a clock on the shuttle, and the spatial coordinates $\xi^{\bar{a}}$ measured along a fixed coordinate grid in the shuttle. This system is clearly not in uniform motion relative to the cosmic rest frame.

Inertial frames are convenient for two reasons:

- In inertial frames the equations of physics assume their pre-GTR form, i.e. without Christoffel symbols. In the ξ coordinates e.g. Maxwell's equations, the conservation law for electric charge, and the Lorentz force law hold up to terms of order $\mathcal{O}(\xi)$ in the form in which you have learned them (written in their special relativistic forms for brevity):

$$\partial_a F^{ab} = -\mu_0 j^b, \quad \partial_a j^a = 0, \quad \ddot{x}^a = \frac{q}{m} F^a{}_b \dot{x}^b, \tag{6.50}$$

  whereas in general they contain Christoffel symbols:

$$\partial_\mu F^{\mu\nu} + \Gamma^\mu{}_{\sigma\mu} F^{\sigma\nu} = -\mu_0 j^\nu, \quad \partial_\mu j^\mu + \Gamma^\mu{}_{\sigma\mu} j^\sigma = 0, \tag{6.51}$$

$$\ddot{x}^\mu + \Gamma^\mu{}_{\rho\nu} \dot{x}^\rho \dot{x}^\nu = \frac{q}{m} F^\mu{}_\nu \dot{x}^\nu. \tag{6.52}$$

- In proofs of local geometric properties of spacetime it is extremely convenient to switch to a local inertial frame. This is a legitimate technique, since geometric properties cannot depend on any particular coordinate system, and therefore one might just as well employ the most convenient system.

We will first demonstrate that local inertial frames can be constructed around a point and then extend the construction along a geodesic curve.

Local inertial frames around a fixed point

In a neighborhood of a fixed point X in spacetime we can always find local coordinates ξ^a such that in these coordinates the metric coefficients in the point X reduce to the Minkowski metric:

$$g_{ab}(X) = \eta_{ab} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{6.53}$$

and the metric does not change in first order if we go away from the point X,

$$\partial_a g_{bc}(X) = 0. \tag{6.54}$$

Coordinates with such properties are denoted as *normal coordinates* in X in the mathematical literature, and they were introduced (for positive definite metrics) by Riemann.

For a construction up to second order in the old coordinates we proceed in the following way: Determine 16 constant coefficients $e_\mu{}^a$ which satisfy the 10 conditions

$$g^{\mu\nu}(X)e_\mu{}^a e_\nu{}^b = \eta^{ab}. \tag{6.55}$$

Writing this equation in matrix form and taking the determinant implies

$$\mathrm{Det}(\underline{e}^T) \cdot \mathrm{Det}(\underline{g}^{-1}) \cdot \mathrm{Det}(\underline{e}) = \mathrm{Det}(\underline{g})^{-1} \cdot \mathrm{Det}(\underline{e})^2 = \mathrm{Det}(\underline{\eta}) = -1, \tag{6.56}$$

and therefore $\mathrm{Det}(\underline{e}) \neq 0$. This implies existence of the inverse matrix $e^\mu{}_a$,

$$e_\mu{}^a e^\mu{}_b = \eta^a{}_b, \quad e_\mu{}^a e^\nu{}_a = g_\mu{}^\nu. \tag{6.57}$$

We can easily check that

$$e^\mu{}_b = g^{\mu\nu} e_\nu{}^c \eta_{bc} \tag{6.58}$$

by calculating e.g.

$$e_\mu{}^a e^\mu{}_b = g^{\mu\nu} e_\mu{}^a e_\nu{}^c \eta_{cb} = \eta^{ac}\eta_{cb} = \eta^a{}_b. \tag{6.59}$$

A frame satisfying the equations (6.55) is denoted as a *tetrad* or a *moving frame*[5].

Instead of the old coordinates x^μ we now introduce new local coordinates ξ^a around X, which up to terms of order $\mathcal{O}[(x-X)^3]$ are given by:

$$\xi^a = e_\mu{}^a(x^\mu - X^\mu) + \frac{1}{2}e_\mu{}^a \Gamma^\mu{}_{\alpha\beta}(X)(x^\alpha - X^\alpha)(x^\beta - X^\beta) \tag{6.60}$$

with inversion (up to terms of order $\mathcal{O}(\xi^3)$)

$$x^\mu - X^\mu = e^\mu{}_a \xi^a - \frac{1}{2}\Gamma^\mu{}_{\alpha\beta}(X) e^\alpha{}_a \xi^a e^\beta{}_b \xi^b. \tag{6.61}$$

Equation (6.60) yields

$$g^{ab}(0) = g^{\mu\nu}(X)\frac{\partial \xi^a}{\partial x^\mu}\bigg|_{x=X} \frac{\partial \xi^b}{\partial x^\nu}\bigg|_{x=X} = g^{\mu\nu}(X) e_\mu{}^a e_\nu{}^b = \eta^{ab}, \tag{6.62}$$

and in first order in $x - X$:

[5] If we do not specify the number of dimensions to be 4, then the tetrad, which is also denoted as a *vierbein* (German origin, literal translation: 'four-leg') is usually called a vielbein (literal translation: 'many-leg'). The French name is *repère mobile*. The names *moving frame* or *repère mobile* are related to the fact that tetrads can be constructed continuously along a geodesic curve through freely falling frames. This is explained below. Tetrads are not only important for the construction of local inertial frames. They are also indispensable for the formulation of equations for spinors in curved spacetime.

$$\begin{aligned} g^{bc}(\xi) &= g^{\alpha\beta}(x)\partial_\alpha\xi^b\partial_\beta\xi^c \\ &= [g^{\alpha\beta}(X) + (x^\rho - X^\rho)\partial_\rho g^{\alpha\beta}(X)][e_\alpha{}^b + e_\kappa{}^b\Gamma^\kappa{}_{\lambda\alpha}(X)(x^\lambda - X^\lambda)] \\ &\quad \times [e_\beta{}^c + e_\sigma{}^c\Gamma^\sigma{}_{\tau\beta}(X)(x^\tau - X^\tau)] \\ &= \eta^{bc} + (x^\rho - X^\rho)[\partial_\rho g^{\alpha\beta}(X) + \Gamma^\alpha{}_{\rho\sigma}(X)g^{\sigma\beta}(X) \\ &\quad + \Gamma^\beta{}_{\rho\sigma}(X)g^{\alpha\sigma}(X)]e_\alpha{}^b e_\beta{}^c \\ &= \eta^{bc} + (x^\rho - X^\rho)\left[D_\rho g^{\alpha\beta}\right]_{x=X} e_\alpha{}^b e_\beta{}^c = \eta^{bc}, \end{aligned} \tag{6.63}$$

because $D_\rho g^{\alpha\beta} = 0$, see equation (5.43). Equation (6.63) implies

$$g^{bc}(\xi) = \eta^{bc} + \mathcal{O}(\xi^2), \tag{6.64}$$

and therefore also

$$[\partial_a g^{bc}(\xi)]_{\xi=0} = 0. \tag{6.65}$$

The results for $g^{ab}|_{\xi=0}$ imply also

$$g_{ab}\,|_{\xi=0} = \eta_{ab},\ [\partial_a g_{bc}(\xi)]_{\xi=0} = -\eta_{bd}\eta_{ce}[\partial_a g^{de}(\xi)]_{\xi=0} = 0, \tag{6.66}$$

and

$$\Gamma^a{}_{bc}|_{\xi=0} = \frac{1}{2}\eta^{ad}(\partial_b g_{cd} + \partial_c g_{bd} - \partial_d g_{bc})\bigg|_{\xi=0} = 0. \tag{6.67}$$

In these coordinates a neighborhood of X looks like a patch of Minkowski spacetime, and the equation of motion for a free (i.e. geodesically moving) particle reduces to

$$\frac{d^2\xi^a}{d\tau^2}\bigg|_{\xi=0} = 0, \tag{6.68}$$

i.e. $t = \xi^0/c \propto \tau +$ const. and the geodesic equation is the free Newton equation:

$$\frac{d^2\boldsymbol{\xi}}{dt^2}\bigg|_{\xi=0} = \mathbf{0}. \tag{6.69}$$

Such a coordinate system defines a *local inertial frame* in the point X.

Freely falling frames

We have seen that a tetrad in the point X,

$$g^{\mu\nu}(X)e_\mu{}^a(X)e_\nu{}^b(X) = \eta^{ab}, \tag{6.70}$$

defines new coordinates ξ^a (6.60) and (6.61) in a neighborhood of X such that the new coordinates define a local inertial frame (6.66) and (6.67) around X. The new basis vectors at X are

$$\epsilon_a \big|_{\xi=0} = \partial_a x^\mu \big|_{\xi=0} e_\mu(X) = e^\mu{}_a e_\mu(X). \tag{6.71}$$

The construction of inertial frames can be extended along a timelike geodesic $\tilde{\gamma} \supset X$ by choosing the unit tangent vector $\dot{x} \equiv dx/d\xi^0$ as the timelike basis vector, where $\xi^0 = s = c\tau$ is the eigentime along the geodesic, $(d\xi^0)^2 = -dx^2$,

$$\epsilon_0 = \dot{x} \quad \Rightarrow \quad \epsilon_0^2 = \dot{x}^2 = -1, \tag{6.72}$$

and parallel transporting also all the remaining vectors along $\tilde{\gamma}$. Since $D_\tau \epsilon_0 = 0$ anyway, we can express this requirement in the form

$$D_\tau \epsilon_a = e_\mu (D_\tau e^\mu{}_a) = e_\mu \left(\frac{d}{d\tau} e^\mu{}_a + \frac{dx^\rho}{d\tau} \Gamma^\mu{}_{\sigma\rho} e^\sigma{}_a \right) = 0, \tag{6.73}$$

i.e.

$$\frac{d}{d\tau} e^\mu{}_a + \frac{dx^\rho}{d\tau} \Gamma^\mu{}_{\sigma\rho} e^\sigma{}_a = \frac{dx^\rho}{d\tau} \left(\partial_\rho e^\mu{}_a + \Gamma^\mu{}_{\sigma\rho} e^\sigma{}_a \right) = 0. \tag{6.74}$$

This construction can be explained as follows: Assume $\tilde{\gamma}(\tau)$ is a timelike curve in spacetime (not necessarily a geodesic), where τ is the eigentime along the curve. Parallel transport along $\tilde{\gamma}$ preserves scalar products

$$\begin{aligned}
\frac{d}{d\tau}(\epsilon_a \cdot \epsilon_b) &= \frac{d}{d\tau}\left(g_{\mu\nu} e^\mu{}_a e^\nu{}_b \right) = e^\mu{}_a e^\nu{}_b \frac{d}{d\tau} g_{\mu\nu} + g_{\mu\nu} e^\nu{}_b \frac{d}{d\tau} e^\mu{}_a + g_{\mu\nu} e^\mu{}_a \frac{d}{d\tau} e^\nu{}_b \\
&= e^\mu{}_a e^\nu{}_b \frac{dx^\rho}{d\tau} \left(\partial_\rho g_{\mu\nu} - g_{\sigma\nu} \Gamma^\sigma{}_{\mu\rho} - g_{\mu\sigma} \Gamma^\sigma{}_{\nu\rho} \right) \\
&= e^\mu{}_a e^\nu{}_b \frac{dx^\rho}{d\tau} D_\rho g_{\mu\nu} = 0.
\end{aligned} \tag{6.75}$$

Therefore parallel transport of a tetrad along $\tilde{\gamma}$ generates tetrads, $\epsilon_a \cdot \epsilon_b = \eta_{ab}$, everywhere on the curve $\tilde{\gamma}$. However, in general this would give us a new local time parameter $t = \xi^0/c$ in each inertial frame along $\tilde{\gamma}$, and in each frame t would have to be small due to corrections of order $\mathcal{O}(t)$ to the local equations of motion $d^2\xi^a/d\tau^2 = 0$ for freely falling particles. To get a continuous set of inertial frames with a continuous time parameter requires $t = \tau$, i.e. $\epsilon_0 \equiv dx/d\xi^0 = c^{-1} dx/d\tau$. The requirement of parallel transport $D_\tau \epsilon_0 = 0$ then implies the geodesic equation $D_\tau \dot{x}(\tau) = 0$, and we see that we can use this construction only along a geodesic.

Inertial frames along a timelike geodesic are denoted as freely falling frames. To determine the basis vectors of freely falling frames in principle requires the construction of a tetrad

$$\epsilon_0 = \dot{x}, \quad \dot{x} \cdot \epsilon_{\bar{a}} = 0, \quad \epsilon_{\bar{a}} \cdot \epsilon_{\bar{b}} = \eta_{\bar{a}\bar{b}} \tag{6.76}$$

in one point X on the geodesic, and then solution of the three equations

$$D_\tau \epsilon_{\bar{a}} = 0 \tag{6.77}$$

to determine the spatial part of the tetrad everywhere on the geodesic.

6.4 Symmetric spaces and conservation laws

We have seen in section 6.2 how the geodesic equation reproduces Newtonian gravity for slowly moving particles in weakly curved spacetimes. In these cases (i.e. when Newtonian gravity applies) we know from the solution of Newton's equation with attractive $1/r^2$ forces that a system of two massive bodies will move on ellipses, parabolas, or hyperbolas around the center of mass of the two-body system[6].

Our next objective is to understand the motion of (not necessarily slow) particles in gravitational fields which are not necessarily weak. To this purpose we will solve the geodesic equation for a particle moving in the 'gravitational field' (i.e. the curved spacetime) of a non-rotating star. However, before attacking this problem we will see that *symmetries of a spacetime (i.e. of the metric) generate conservation laws.*

Conservation laws in curved spacetime or in curvilinear coordinates

We have seen in section 4.4 that in inertial frames, conservation of a charge Q corresponds to a local conservation law $\partial_\mu j^\mu = 0$. However, in a curved spacetime generically this equation can hold in this form only along a geodesic in a freely falling frame, and the correct general form is $D_\mu j^\mu = 0$. Our objective here is to find the relation between $D_\mu j^\mu = 0$ and conserved charges, and how this generalizes to the stress–energy tensor and momentum conservation in curved spacetimes.

The four-dimensional volume measure in spacetime is[7] $d^4x\sqrt{-g}$, where g is the determinant of the metric. The key observation for the derivation of conserved charges from $D_\mu j^\mu = 0$ is that the derivative of $\sqrt{-g}$ is related to a Christoffel symbol:

$$\partial_\mu\sqrt{-g} = -\frac{1}{2\sqrt{-g}}\frac{\partial g}{\partial g_{\rho\sigma}}\partial_\mu g_{\rho\sigma} = \frac{1}{2}\sqrt{-g}\,g^{\rho\sigma}\partial_\mu g_{\rho\sigma} = \sqrt{-g}\,\Gamma^\sigma{}_{\sigma\mu}, \tag{6.78}$$

and this implies the following general relation between ordinary and covariant divergences:

$$\partial_\mu(\sqrt{-g}\,j^\mu) = \sqrt{-g}\,(\partial_\mu j^\mu + \Gamma^\mu{}_{\sigma\mu} j^\sigma) = \sqrt{-g}\,D_\mu j^\mu. \tag{6.79}$$

Integrating $\partial_\mu(\sqrt{-g}\,j^\mu) = 0$ from t_0 to t and over a three-dimensional fixed volume V yields a conservation law for the charge

$$Q_V = \frac{1}{c}\int_V d^3\boldsymbol{x}\sqrt{-g}\,j^0 \tag{6.80}$$

[6] See e.g. Goldstein H, Poole C and Safko J 2002 *Classical Mechanics* 3rd edn (San Francisco, CA: Addison Wesley).

[7] We can use our knowledge that there exist local inertial frames in every point X to prove this: In the local inertial coordinates the volume measure in the point X is $d^4\xi$, and the inertial coordinates are related to arbitrary coordinates with metric $g_{\mu\nu}$ through $\partial_\mu\xi^a\,\partial_\nu\xi^b\,\eta_{ab} = g_{\mu\nu}$. This implies for the determinants $-\mathrm{Det}(\partial_\mu\xi^a)^2 = g$, and therefore $d^4\xi = d^4x\,|\mathrm{Det}(\partial_\mu\xi^a)| = d^4x\sqrt{-g}$.

in the volume:

$$
\begin{aligned}
Q_V(t) - Q_V(t_0) &= \frac{1}{c}\int_V d^3\boldsymbol{x}\sqrt{-g}j^0\,\Big|_{t_0}^{t} = \int_{t_0}^{t} dt'\int_V d^3\boldsymbol{x}\;\partial_0(\sqrt{-g}j^0) \\
&= -\int_{t_0}^{t} dt'\int_V d^3\boldsymbol{x}\;\nabla\cdot(\sqrt{-g}\boldsymbol{j}) = -\int_{t_0}^{t} dt'\oint_{\partial V} d^2S\;\boldsymbol{n}_S\cdot\sqrt{-g}\boldsymbol{j},
\end{aligned}
\tag{6.81}
$$

i.e.

$$\frac{d}{dt}Q_V = -\oint_{\partial V} d^2S\;\boldsymbol{n}_S\cdot\sqrt{-g}\boldsymbol{j}. \tag{6.82}$$

The Killing equation and Killing vectors

Under coordinate transformations $x \to x'(x)$ the metric transforms covariantly

$$g_{\mu\nu}(x) \to g'_{\mu\nu}(x') = \partial'_\mu x^\alpha \partial'_\nu x^\beta\; g_{\alpha\beta}(x), \tag{6.83}$$

or

$$g_{\alpha\beta}(x) = \partial_\alpha x'^\mu\;\partial_\beta x'^\nu\; g'_{\mu\nu}(x'). \tag{6.84}$$

It does not matter for these equations whether the transformation is only a passive coordinate transformation (i.e. just a relabeling of the same point in spacetime with a new set of coordinates), or whether it is an active transformation from the initial point with coordinates x to the final point with coordinates x'. The figures 2.2 and 2.3 explain the difference between passive and active transformations in the case of translations. In either case the transformation $x \to x'(x)$ is a *symmetry* of the metric if the new metric $g'_{\mu\nu}(x')$ is *exactly the same function* of its arguments x'^μ as the old metric $g_{\mu\nu}(x')$, i.e. if we have

$$g'_{\mu\nu}(x') = g_{\mu\nu}(x'). \tag{6.85}$$

This property is denoted as *form invariance* of the metric under the transformation $x \to x'(x)$.

Equations (6.84) and (6.85) imply for a symmetry transformation $x \to x'(x)$ the condition

$$g_{\alpha\beta}(x) = \partial_\alpha x'^\mu\;\partial_\beta x'^\nu\; g_{\mu\nu}(x'). \tag{6.86}$$

We will evaluate this in first order in the coordinate shifts $\epsilon^\mu(x)$,

$$x'^\mu(x) = x^\mu + \delta x^\mu(x) = x^\mu - \epsilon^\mu(x). \tag{6.87}$$

The minus sign in this equation is a matter of convention and is motivated from the passive interpretation of coordinate translations. For example if we move the origin of our coordinate system x^μ by an amount ϵ^1 in the x^1 direction, the new x'^1 coordinate of the point with old coordinates x^1 is $x'^1 = x^1 - \epsilon^1$, see figure 2.2.

First-order evaluation of the equation

$$g_{\alpha\beta}(x) = \partial_\alpha[x^\mu - \epsilon^\mu(x)]\,\partial_\beta[x^\nu - \epsilon^\nu(x)]\,g_{\mu\nu}(x - \epsilon(x)) \tag{6.88}$$

yields

$$\begin{aligned}
0 &= \partial_\alpha\epsilon^\mu(x)g_{\mu\beta}(x) + \partial_\beta\epsilon^\mu(x)g_{\alpha\mu}(x) + \epsilon^\mu(x)\partial_\mu g_{\alpha\beta}(x)\\
&= \partial_\alpha\Big(\epsilon^\mu(x)g_{\mu\beta}(x)\Big) + \partial_\beta\Big(\epsilon^\mu(x)g_{\alpha\mu}(x)\Big)\\
&\quad + \epsilon^\mu(x)\Big(\partial_\mu g_{\alpha\beta}(x) - \partial_\alpha g_{\mu\beta}(x) - \partial_\beta g_{\alpha\mu}(x)\Big)\\
&= \partial_\alpha\epsilon_\beta(x) + \partial_\beta\epsilon_\alpha(x) - 2\epsilon_\mu(x)\Gamma^\mu{}_{\alpha\beta}(x) = D_\alpha\epsilon_\beta(x) + D_\beta\epsilon_\alpha(x).
\end{aligned}$$

We see that an infinitesimal coordinate displacement $\epsilon^\mu(x)$ has to satisfy the *Killing equation*[8]

$$D_\mu\epsilon_\nu(x) + D_\nu\epsilon_\mu(x) = 0 \tag{6.89}$$

to generate a symmetry. The solutions $\epsilon_\mu(x)$ of the Killing equation are denoted as *Killing vectors*.

Conservation laws for particle motion

We have seen that the equation of motion of a free particle in terms of its eigentime τ is

$$\ddot{x}^\mu + \Gamma^\mu{}_{\rho\nu}\dot{x}^\rho\dot{x}^\nu = 0. \tag{6.90}$$

We can now easily demonstrate that the geodesic equation (6.90) and the Killing equation (6.89) together imply the conservation law

$$\frac{d}{d\tau}(\epsilon_\mu\dot{x}^\mu) = 0. \tag{6.91}$$

For the proof we observe

$$\begin{aligned}
\frac{d}{d\tau}(\epsilon_\mu\dot{x}^\mu) &= \dot{x}^\rho\frac{d}{d\tau}\epsilon_\rho - \epsilon_\mu\Gamma^\mu{}_{\rho\nu}\dot{x}^\rho\dot{x}^\nu = \dot{x}^\rho\dot{x}^\nu(\partial_\nu\epsilon_\rho - \epsilon_\mu\Gamma^\mu{}_{\rho\nu}) = \dot{x}^\rho\dot{x}^\nu D_\nu\epsilon_\rho\\
&= \frac{1}{2}\dot{x}^\rho\dot{x}^\nu(D_\nu\epsilon_\rho + D_\rho\epsilon_\nu) = 0.
\end{aligned} \tag{6.92}$$

Time translation invariance and energy conservation

Suppose the metric of a spacetime has the property to be time-independent,

$$g_{\mu\nu}(x) = g_{\mu\nu}(\boldsymbol{x}), \quad \partial_0 g_{\mu\nu}(x) = 0. \tag{6.93}$$

[8] Killing W 1892 *J. Reine Angew. Math.* **109** 121.

Then this metric is apparently invariant under constant shifts of the time coordinate $x^0 = ct$,

$$x'^\mu = x^\mu + g^\mu{}_0 \delta x^0, \quad \partial_\alpha \delta x^0 = 0, \tag{6.94}$$

$$\partial_\alpha x'^\mu = \partial_\alpha x^\mu = g_\alpha{}^\mu. \tag{6.95}$$

This implies

$$g_{\alpha\beta}(x) = \partial_\alpha x'^\mu \, \partial_\beta x'^\nu \, g'_{\mu\nu}(x') = g'_{\alpha\beta}(x'), \tag{6.96}$$

and since there is no time-dependence, and $\boldsymbol{x}' = \boldsymbol{x}$, the previous equation just says that the new metric components depend on their arguments in exactly the same way as the old metric components,

$$g_{\alpha\beta}(\boldsymbol{x}) = g'_{\alpha\beta}(\boldsymbol{x}). \tag{6.97}$$

Therefore, the vector

$$\epsilon_\nu = g_{\nu\mu}\epsilon^\mu = -g_{\nu 0}\delta x^0 \tag{6.98}$$

with constant δx^0 must satisfy the Killing equation in a spacetime with metric $g_{\alpha\beta}(\boldsymbol{x})$. Let us check explicitly that this is true. We have

$$\begin{aligned} D_\mu \epsilon_\nu &= \partial_\mu \epsilon_\nu - \epsilon_\rho \Gamma^\rho{}_{\nu\mu} \\ &= -\delta x^0 \partial_\mu g_{\nu 0} + \delta x^0 g_{\rho 0} \frac{1}{2} g^{\rho\sigma}\left(\partial_\mu g_{\sigma\nu} + \partial_\nu g_{\sigma\mu} - \partial_\sigma g_{\mu\nu}\right) \\ &= -\delta x^0 \partial_\mu g_{\nu 0} + \frac{1}{2}\delta x^0 (\partial_\mu g_{0\nu} + \partial_\nu g_{0\mu} - \partial_0 g_{\mu\nu}) \\ &= -\frac{1}{2}\delta x^0 (\partial_\mu g_{0\nu} - \partial_\nu g_{0\mu}), \end{aligned} \tag{6.99}$$

and therefore we find indeed that the Killing equation (6.89) is satisfied. The corresponding conserved quantity

$$\epsilon_\mu \dot{x}^\mu = -\delta x^0 g_{0\mu}\dot{x}^\mu = -\frac{\delta x^0}{m} p_0 = \frac{\delta x^0}{mc} E \tag{6.100}$$

is up to the constant factor $\delta x^0/mc$ just the energy of the particle. Free particles moving in a *time-independent* spacetime have conserved energy, but free particles moving in a *time-dependent* spacetime (like the expanding Universe) do *not* have conserved energy. This observation also holds for more complicated physical systems involving electromagnetic fields and quantum fields: The energy of physical systems is conserved in *time-independent* spacetimes but *not* in *evolving* spacetimes.

Spatial translation invariance and momentum conservation

Suppose the metric components do not depend on the particular spatial coordinate x^i. This implies a symmetry of the metric under

$$x^\mu \to x'^\mu = x^\mu - \epsilon g_i{}^\mu \tag{6.101}$$

with constant parameter ϵ. Note that the translation (6.101) satisfies the equation $\partial_\alpha \delta x^\mu = 0$ because both factors in $\delta x^\mu = -\epsilon g_i{}^\mu$ do not depend on the position x in spacetime. This implies

$$g_{\alpha\beta}(x) = \partial_\alpha x'^\mu \, \partial_\beta x'^\nu \, g'_{\mu\nu}(x') = g'_{\alpha\beta}(x') = g_{\alpha\beta}(x'), \tag{6.102}$$

since by assumption, $g_{\alpha\beta}(x)$ does not depend on the only shifted coordinate x^i and therefore $g_{\alpha\beta}(x) = g_{\alpha\beta}(x')$. The symmetry condition $g_{\alpha\beta}(x) = g'_{\alpha\beta}(x)$ is therefore fulfilled.

We can also verify that the Killing equation holds. Equation (5.42) implies

$$D_\mu \epsilon_\nu = \partial_\mu \epsilon_\nu - \epsilon_\rho \Gamma^\rho{}_{\nu\mu} = \epsilon(\partial_\mu g_{i\nu} - g_{i\rho}\Gamma^\rho{}_{\nu\mu}) = \epsilon g_{\rho\nu}\Gamma^\rho{}_{i\mu}, \tag{6.103}$$

and substitution of equation (5.13) then confirms with $\partial_i g_{\nu\mu} = 0$ the Killing equation,

$$D_\mu \epsilon_\nu = \frac{\epsilon}{2}(\partial_i g_{\nu\mu} + \partial_\mu g_{i\nu} - \partial_\nu g_{i\mu}) = \frac{\epsilon}{2}(\partial_\mu g_{i\nu} - \partial_\nu g_{i\mu}) = -D_\nu \epsilon_\mu. \tag{6.104}$$

The corresponding conserved quantity $\epsilon_\mu \dot{x}^\mu$ is after division by the irrelevant constant factor ϵ the covariant spatial component $u_i(\tau)$ of the 4-velocity of the particle,

$$g_{i\mu}(x(\tau))\dot{x}^\mu(\tau) = u_i(\tau) = p_i(\tau)/m. \tag{6.105}$$

Spatial translation invariance of the metric in the direction of the coordinate x^i implies conservation of the spatial momentum component $p_i = mu_i$.

Rotational invariance and angular momentum conservation

Suppose the metric $ds^2 = g_{\mu\nu}(x)dx^\mu dx^\nu$ depends on the three-dimensional spatial coordinates x^i only through the products

$$r^2 \equiv \sum_{i=1}^{3}(x^i)^2, \quad rdr = \sum_{i=1}^{3} x^i dx^i. \tag{6.106}$$

Then the metric must have the form[9]

$$\begin{aligned} ds^2 = g_{00}(x^0, r)dx^0dx^0 + 2H(x^0, r)dx^0 \sum_{i=1}^{3} x^i dx^i + J(x^0, r)\sum_{i=1}^{3}(dx^i)^2 \\ + K(x^0, r)\sum_{i,j=1}^{3} x^i x^j dx^i dx^j, \end{aligned} \tag{6.107}$$

[9] Below I will remark that this metric can be further simplified (namely to $H = 0$ and $J = 1$) through appropriate coordinate transformations. We will see that explicitly in section 7.3, when we derive the metric outside a non-rotating spherical mass. Using these simplifications would make the calculations in this section shorter. However, the easy and most elegant way is not always the most instructive way.

with

$$g_{0i}(x) = H(x^0, r)x^i, \quad g_{ij}(x) = J(x^0, r)\delta_{ij} + K(x^0, r)x^i x^j. \tag{6.108}$$

Note the non-covariant notation, since we singled out certain classes of coordinates where the metric has this special form. In particular we will not draw any indices with the metric until we reach equation (6.120).

The metric components in equation (6.107) are form invariant under rotations

$$x^i \to x'^i = R^i{}_j x^j \tag{6.109}$$

of the spatial coordinates, where $\underline{R}$ is a standard 3×3 rotation matrix. Let us check this invariance explicitly. We have

$$\partial_0 x'^0 = 1, \quad \partial_i x'^k = R^k{}_i, \tag{6.110}$$

and all other components of the Jacobian matrix vanish. Therefore the old metric components in terms of the new components are (with $r' = r$ under rotations)

$$g_{00}(x^0, r) = g'_{00}(x') = g'_{00}(x'^0, r) = g'_{00}(x^0, r), \tag{6.111}$$

$$g_{0i}(x) = H(x^0, r)x^i = \partial_i x'^k g'_{0k}(x') = R^k{}_i g'_{0k}(x') \tag{6.112}$$

$$\begin{aligned} \Rightarrow g'_{0k}(x') &= H(x^0, r)x^i (R^{-1})^i{}_k = H(x^0, r)x'^k = H(x'^0, r)x'^k \\ &= g_{0k}(x'), \end{aligned} \tag{6.113}$$

and finally

$$\begin{aligned} g_{ij}(x) &= J(x^0, r)\delta_{ij} + K(x^0, r)x^i x^j = \partial_i x'^m \partial_j x'^n g'_{mn}(x') \\ &= R^m{}_i R^n{}_j g'_{mn}(x'), \end{aligned} \tag{6.114}$$

which implies

$$\begin{aligned} g'_{mn}(x') &= [J(x^0, r)\delta_{ij} + K(x^0, r)x^i x^j](R^{-1})^i{}_m (R^{-1})^j{}_n \\ &= J(x^0, r)\delta_{mn} + K(x^0, r)x'^m x'^n \\ &= J(x'^0, r')\delta_{mn} + K(x'^0, r')x'^m x'^n = g_{mn}(x'), \end{aligned} \tag{6.115}$$

i.e. the components of the metric (6.107) are really form invariant under the rotations (6.109).

In first order a three-dimensional rotation by a small angle can be written in the form

$$\boldsymbol{x}' = \underline{R} \cdot \boldsymbol{x} \simeq \boldsymbol{x} - \boldsymbol{\Phi} \times \boldsymbol{x}, \tag{6.116}$$

or in components

$$x'^i = R^i{}_j x^j = x^i - \epsilon_{ijk}\Phi^j x^k \tag{6.117}$$

where the direction of the constant vector $\boldsymbol{\Phi}$ is the rotation axis of the coordinate system in passive interpretation (or the direction of $-\boldsymbol{\Phi}$ is the rotation axis in active interpretation), and the magnitude

$$|\mathbf{\Phi}| \equiv \sqrt{\sum_i (\Phi^i)^2} \ll 1 \tag{6.118}$$

is the small rotation angle. Note again the non-covariant form of the equation (6.117), which reflects the fact that we have broken general covariance by singling out special coordinate systems where the metric has the manifestly rotation symmetric form (6.107).

According to equations (6.116) and (6.117), the Killing vectors generating the symmetry transformations (6.109) are

$$\boldsymbol{\epsilon} = \mathbf{\Phi} \times \boldsymbol{x}, \quad \epsilon^i = \epsilon_{ijk}\Phi^j x^k. \tag{6.119}$$

The corresponding covariant components $\epsilon_\mu = g_{\mu\nu}\epsilon^\nu = g_{\mu i}\epsilon^i$ of the Killing vectors are

$$\epsilon_0 = H\sum_{i=1}^{3} x^i\epsilon^i = 0, \quad \epsilon_i = J\epsilon^i + Kx^i\sum_{j=1}^{3} x^j\epsilon^j = J\epsilon^i = -J\Phi^j\epsilon_{jik}x^k. \tag{6.120}$$

We can also think of the Killing vector $\boldsymbol{\epsilon}$ as a linear combination of three independent Killing vectors,

$$\boldsymbol{\epsilon} = \Phi^i\boldsymbol{\epsilon}_{(i)} \tag{6.121}$$

with the components $\epsilon_{(i)j}$ of the vector $\boldsymbol{\epsilon}_{(i)}$ following from equation (6.120),

$$\epsilon_{(i)j}(x^0, r) = J(x^0, r)\epsilon_{(i)}{}^j = -J(x^0, r)\epsilon_{ijk}x^k. \tag{6.122}$$

The corresponding conserved quantities involve the four-dimensional scalar product $\epsilon_\mu\dot{x}^\mu = \epsilon^\nu g_{\nu\mu}\dot{x}^\mu$, and therefore we now have to be careful about index positions. The conserved quantity

$$\epsilon_\mu\dot{x}^\mu = \epsilon^i g_{i\mu}(x)\dot{x}^\mu = \epsilon^i u_i = (\mathbf{\Phi} \times \boldsymbol{x}) \cdot \boldsymbol{u} = \mathbf{\Phi} \cdot (\boldsymbol{x} \times \boldsymbol{u}) \tag{6.123}$$

contains three independent transformation parameters Φ^i and therefore yields a conserved 'vector' (actually a triplet of conserved quantities that looks like a vector from a three-dimensional point of view):

$$\boldsymbol{L} = m\boldsymbol{x} \times \boldsymbol{u} = \boldsymbol{x} \times \boldsymbol{p}. \tag{6.124}$$

Here $u_i = g_{i\mu}\dot{x}^\mu$ are the spatial components of the 4-velocity. $\boldsymbol{L}$ is the conserved angular momentum of the particle moving through the rotationally symmetric spacetime.

Note that the following technical remarks apply to the angular momentum of the particle:

- The three-dimensional products of the form $\boldsymbol{a} \cdot \boldsymbol{b}$ in equation (6.123) correspond to $\boldsymbol{a} \cdot \boldsymbol{b} = \sum_{i=1}^{3} a^i b_i$, with exactly these index positions in the first and second factor.
- In equation (6.124) the vector $\boldsymbol{x}$ is the three-dimensional vector with components x^i, but the vector $\boldsymbol{u}$ has components $u_i = g_{i\mu}\dot{x}^\mu = g_{i\mu}dx^\mu/d\tau$.

- The triplet $\boldsymbol{L}$ of conserved angular momentum components for motion in a radially symmetric metric is *not* the spatial part of a 4-vector.
- Conservation of the angular momentum $\boldsymbol{L}$ in a rotationally symmetric spacetime holds irrespective of whether any of the 4-velocity components $u_i = p_i/m$ is conserved, i.e. there is *no presumption* of translation invariance with respect to constant shifts of any of the spatial coordinates x^i. Indeed, a completely rotationally symmetric metric which is also invariant under any particular spatial translation would have to be invariant under all spatial translations, i.e. be a purely time-dependent metric. However, we can have the case that the metric is only invariant under rotations around the x^i direction. Then it *may* also be invariant under translations along the x^i direction (cylinder symmetry). In that case the two quantities $L_i = \epsilon_{ijk}x^j p_k$ and p_i would be conserved. Or it *may* be the case that the metric is invariant under translations perpendicular to the rotation axis x^i. Then the angular momentum component L_i and the two perpendicular momentum components $P_j = \epsilon_{ijk}p_k$ would be conserved.

We can verify the conservation laws for the components of $\boldsymbol{L}$ explicitly. The geodesic equation for the metric (6.107) yields for $\dot{u}_i = du_i/d\tau$ (with $\partial_i r = x^i/r$):

$$\begin{aligned}\dot{u}_i &= \frac{1}{2}\dot{x}^\mu \dot{x}^\nu \partial_i g_{\mu\nu} \\ &= \left(\frac{1}{2}(\dot{x}^0)^2\partial_r g_{00} + \dot{x}^0\sum_k x^k\dot{x}^k\partial_r H + \frac{1}{2}\sum_k(\dot{x}^k)^2\partial_r J \right. \\ &\quad \left. + \frac{1}{2}\sum_{m,n} x^m x^n \dot{x}^m \dot{x}^n \partial_r K\right)\frac{x^i}{r} + \left(H\dot{x}^0 + K\sum_k x^k\dot{x}^k\right)\dot{x}^i.\end{aligned}$$

This yields for the components $x^i u_j - x^j u_i$ of $\boldsymbol{L}/m$:

$$\begin{aligned}\frac{d}{d\tau}\left(x^i u_j - x^j u_i\right) &= \dot{x}^i u_j - \dot{x}^j u_i + x^i \dot{u}_j - x^j \dot{u}_i \\ &= \dot{x}^i g_{j\mu}\dot{x}^\mu - \dot{x}^j g_{i\mu}\dot{x}^\mu + \left(H\dot{x}^0 + K\sum_k x^k\dot{x}^k\right) \\ &\quad \times (x^i\dot{x}^j - x^j\dot{x}^i) \\ &= H\dot{x}^0(\dot{x}^i x^j - \dot{x}^j x^i) + K\sum_k x^k\dot{x}^k(\dot{x}^i x^j - \dot{x}^j x^i) \\ &\quad + \left(H\dot{x}^0 + K\sum_k x^k\dot{x}^k\right)(x^i\dot{x}^j - x^j\dot{x}^i) = 0,\end{aligned}$$

i.e. the components of $\boldsymbol{L}$ are indeed conserved. Furthermore, if we substitute

$$u_i = g_{i\mu}\dot{x}^\mu = Hx^i\dot{x}^0 + J\dot{x}^i + Kx^i\sum_k x^k\dot{x}^k \tag{6.125}$$

into the components of $\boldsymbol{L}/m$ we find

$$x^i u_j - x^j u_i = J(x^0, r)(x^i\dot{x}^j - x^j\dot{x}^i), \tag{6.126}$$

which indicates that the components of $\boldsymbol{L}$ will actually be part of an antisymmetric tensor.

We will later prove that the metric (6.107) can be simplified to $H(x^0, r) = 0$, $J(x^0, r) = 1$ through appropriate coordinate transformations. Therefore, in these coordinates the conserved angular momentum of a particle moving through a radially symmetric spacetime has components

$$M^{ij} = m[x^i(\tau)\dot{x}^j(\tau) - x^j(\tau)\dot{x}^i(\tau)]. \tag{6.127}$$

Conservation laws for fields and fluids

We have already seen that symmetry of a spacetime under shifts along Killing vectors ϵ^μ implies conservation laws for energy or momentum or angular momentum of freely falling particles. This property also holds for fluids or fields: The stress–energy tensor yields a conserved current $T^\nu_{(\epsilon)} = \epsilon_\mu T^{\mu\nu}$ if combined with a Killing vector ϵ_μ. To elucidate this, we first note that the curvilinear generalization of the local conservation law (4.94) is

$$D_\nu T^{\mu\nu} = 0. \tag{6.128}$$

This equation, the Killing equation (6.89), and the symmetry $T^{\mu\nu} = T^{\nu\mu}$ imply

$$\partial_\nu(\sqrt{-g}\,\epsilon_\mu T^{\mu\nu}) = \sqrt{-g}\,D_\nu(\epsilon_\mu T^{\mu\nu}) = \sqrt{-g}\,T^{\mu\nu}D_\nu\epsilon_\mu = 0. \tag{6.129}$$

The resulting conserved charge is

$$p_{(\epsilon)} = \frac{1}{c}\int_V d^3\boldsymbol{x}\sqrt{-g}\,\epsilon_\mu T^{\mu 0}, \tag{6.130}$$

$$\frac{d}{dt}p_{(\epsilon)} = -\oint_{\partial V} d^2S\,\boldsymbol{n}_S\cdot\sqrt{-g}\,\boldsymbol{T}_{(\epsilon)}, \tag{6.131}$$

where the current density of the momentum $p_{(\epsilon)}$ is

$$\boldsymbol{T}_{(\epsilon)} = \epsilon_\mu T^{\mu i}\boldsymbol{e}_i. \tag{6.132}$$

These conserved charges are again energy or momentum or angular momentum if the Killing vectors correspond to timelike or spacelike shifts or spatial rotations, respectively. In these cases we have Killing vectors

$$\epsilon_{(\nu)}{}^\mu = g_\nu{}^\mu \tag{6.133}$$

and corresponding conserved charges

$$p_\nu = \frac{1}{c}\int_V d^3\mathbf{x}\sqrt{-g}\,\epsilon_{(\nu)\mu}T^{\mu 0} = \frac{1}{c}\int_V d^3\mathbf{x}\sqrt{-g}\,T_\nu{}^0. \tag{6.134}$$

This is consistent with the corresponding results for single particle momenta: The conserved momentum for $\partial L/\partial x^\mu = 0$ is

$$p_\mu = \frac{\partial L}{\partial \dot{x}^\mu} = mg_{\mu\nu}\dot{x}^\nu = m\epsilon_{(\mu)}\cdot\dot{x}. \tag{6.135}$$

We note in particular that the stress–energy tensor of electromagnetic fields in curvilinear coordinates or in a curved spacetime is

$$T^{\mu\nu} = \frac{1}{\mu_0}\left(F^{\mu\lambda}F^\nu{}_\lambda - \frac{1}{4}g^{\mu\nu}F_{\kappa\lambda}F^{\kappa\lambda}\right), \tag{6.136}$$

and if there are no charges or currents this tensor satisfies

$$D_\nu T^{\mu\nu} = \partial_\nu T^{\mu\nu} + \Gamma^\mu{}_{\sigma\nu}T^{\sigma\nu} + \Gamma^\nu{}_{\sigma\nu}T^{\mu\sigma} = 0. \tag{6.137}$$

This implies conservation laws for electromagnetic energy or momentum in the directions of translation symmetries in spacetimes.

Corresponding remarks apply to the curvilinear generalization

$$T^{\mu\nu} = (\varrho + p)u^\mu u^\nu + pg^{\mu\nu} \tag{6.138}$$

of the stress–energy tensor (4.114) of a perfect fluid.

IOP Concise Physics

Chapter 7

The dynamics of spacetime: the Einstein equation

We have discussed equations of motion for particles and fields in curved spacetimes or in curvilinear coordinates, but we have not yet figured out the equations which actually govern the dynamics of spacetime itself. We will fill this gap in this chapter.

7.1 Geodesic deviation and curvature

Consider two freely falling observers who fall along different geodesics $x(\tau)$ and $x(\tau) + \delta x(\tau)$. Both observers have their own eigentimes, but we want to compare locations of the two observers when their two eigentimes have the same value τ. We also assume that the two observers are close together at equal values τ of their eigentimes, $|\delta x^\mu(\tau)| \ll |x^\mu(\tau)|$.

If we expand the geodesic equation for $x(\tau) + \delta x(\tau)$ in first order in $\delta x(\tau)$ and take into account the geodesic equation for $x(\tau)$, we find

$$\frac{d^2\delta x^\mu}{d\tau^2} + 2\Gamma^\mu{}_{\nu\rho}(x)\frac{dx^\nu}{d\tau}\frac{d\delta x^\rho}{d\tau} + \delta x^\rho \partial_\rho \Gamma^\mu{}_{\nu\rho}(x)\frac{dx^\nu}{d\tau}\frac{dx^\sigma}{d\tau} = 0. \tag{7.1}$$

However, coordinate shifts δx^μ transform like contravariant vectors under coordinate transformations:

$$\delta x^\alpha \to \delta x'^\mu = x'^\mu(x + \delta x) - x'^\mu(x) = \delta x^\alpha \partial_\alpha x'^\mu, \tag{7.2}$$

and therefore they have to be differentiated like contravariant vectors,

$$D_\tau \delta x^\mu = \frac{dx^\nu}{d\tau} D_\nu \delta x^\mu = \frac{d\delta x^\mu}{d\tau} + \Gamma^\mu{}_{\rho\nu}\delta x^\rho \frac{dx^\nu}{d\tau}. \tag{7.3}$$

In the language of the projectors from section 5.2: The coordinate shifts δx correspond to a tangential shift at the point x, but they need not correspond to a

doi:10.1088/2053-2571/aaf173ch7

tangential shift at an adjacent point, and therefore one has to invoke the local projectors $\underline{P}$ from section 5.2 to define the derivative of δx.

The second derivative is then

$$\begin{aligned} D_\tau^2\delta x^\mu &= \frac{d}{d\tau}\left(\frac{d\delta x^\mu}{d\tau} + \Gamma^\mu{}_{\rho\nu}\delta x^\rho\frac{dx^\nu}{d\tau}\right) + \Gamma^\mu{}_{\lambda\sigma}\left(\frac{d\delta x^\lambda}{d\tau} + \Gamma^\lambda{}_{\rho\nu}\delta x^\rho\frac{dx^\nu}{d\tau}\right)\frac{dx^\sigma}{d\tau} \\ &= \frac{d^2\delta x^\mu}{d\tau^2} + \partial_\sigma\Gamma^\mu{}_{\rho\nu}\delta x^\rho\frac{dx^\sigma}{d\tau}\frac{dx^\nu}{d\tau} + \Gamma^\mu{}_{\rho\nu}\delta x^\rho\frac{d^2x^\nu}{d\tau^2} + 2\Gamma^\mu{}_{\rho\nu}\frac{d\delta x^\rho}{d\tau}\frac{dx^\nu}{d\tau} \\ &\quad + \Gamma^\mu{}_{\lambda\sigma}\Gamma^\lambda{}_{\rho\nu}\frac{dx^\sigma}{d\tau}\delta x^\rho\frac{dx^\nu}{d\tau}. \end{aligned} \tag{7.4}$$

Substitution of equation (7.1) for $d^2\delta x^\mu/d\tau^2$ and the geodesic equation for $d^2x^\nu/d\tau^2$ yields the *equation of geodesic deviation*

$$\begin{aligned} D_\tau^2\delta x^\mu &= \delta x^\rho\frac{dx^\sigma}{d\tau}\frac{dx^\nu}{d\tau}\left(\partial_\sigma\Gamma^\mu{}_{\rho\nu} - \partial_\rho\Gamma^\mu{}_{\sigma\nu} + \Gamma^\mu{}_{\lambda\sigma}\Gamma^\lambda{}_{\rho\nu} - \Gamma^\mu{}_{\rho\lambda}\Gamma^\lambda{}_{\sigma\nu}\right) \\ &= -\frac{dx^\nu}{d\tau}\delta x^\rho\frac{dx^\sigma}{d\tau}R^\mu{}_{\nu\rho\sigma}, \end{aligned} \tag{7.5}$$

with the *Riemann curvature tensor*

$$R^\mu{}_{\nu\rho\sigma} = \partial_\rho\Gamma^\mu{}_{\nu\sigma} - \partial_\sigma\Gamma^\mu{}_{\nu\rho} + \Gamma^\mu{}_{\lambda\rho}\Gamma^\lambda{}_{\nu\sigma} - \Gamma^\mu{}_{\lambda\sigma}\Gamma^\lambda{}_{\nu\rho}. \tag{7.6}$$

Equation (7.5) is an equation for the covariant acceleration $D_\tau^2\delta x$ of the separation δx between adjacent freely falling particles. Therefore the right-hand side of this equation can be interpreted as the local *tidal force* per mass of a freely falling particle. This becomes apparent in a local inertial frame:

$$\frac{d^2\delta\xi^a}{d\tau^2} = -\frac{d\xi^b}{d\tau}\delta\xi^c\frac{d\xi^d}{d\tau}R^a{}_{bcd} = -c^2\delta\xi^c R^a{}_{0c0}. \tag{7.7}$$

Here the covariant derivatives on the left-hand side become ordinary derivatives due to $\Gamma^a{}_{bc} = 0$ in the origin of the local inertial frame, and the property $\xi^0 = c\tau \Rightarrow d\xi^b/d\tau = c\eta^b{}_0$ was used.

In physics terms, curvature means that the rate of change $\partial_\tau\delta\boldsymbol{\xi}$ of the distance vectors between freely falling particles is not constant: $\partial_\tau^2\delta\boldsymbol{\xi} \neq 0$. Formulated in another way: Spacetime is curved if freely falling particles do not move with constant velocities in the rest frame of one of the freely falling particles.

The tidal force is a measure of how strong the local curvature of spacetime pushes geodesics together or pulls them apart, and it is the tool to actually physically distinguish between curved spacetimes on the one hand and the use of curvilinear coordinates in Minkowski spacetime on the other hand. There is curvature if and only if we can observe tidal forces affecting the relative motions of free particles.

Symmetries of the Riemann tensor

The symmetries of the Riemann tensor are most easily revealed in a local inertial frame: $\Gamma^a{}_{bc} = 0$ in the point X implies for the Riemann tensor in that point

$$\begin{aligned} R_{abcd} &= \frac{1}{2}\partial_c(\partial_b g_{ad} + \partial_d g_{ab} - \partial_a g_{bd}) - \frac{1}{2}\partial_d(\partial_b g_{ac} + \partial_c g_{ab} - \partial_a g_{bc}) \\ &= \frac{1}{2}(\partial_c\partial_b g_{ad} + \partial_d\partial_a g_{bc} - \partial_c\partial_a g_{bd} - \partial_d\partial_b g_{ac}), \end{aligned} \tag{7.8}$$

which yields in any coordinate system

$$R_{\rho\sigma\mu\nu} = e_\rho{}^a e_\sigma{}^b e_\mu{}^c e_\nu{}^d R_{abcd} = -R_{\sigma\rho\mu\nu} = -R_{\rho\sigma\nu\mu} = R_{\mu\nu\rho\sigma}. \tag{7.9}$$

These equations imply that the Riemann tensor has at most as many independent components as a symmetric (6×6)-matrix, i.e. at most 21 independent components. However, there is one more algebraic identity:

$$R_{\rho\sigma\mu\nu} + R_{\rho\mu\nu\sigma} + R_{\rho\nu\sigma\mu} = 0. \tag{7.10}$$

This equation is already a consequence of equation (7.9) if any two indices coincide. This and the antisymmetry in the first pair of indices imply that equation (7.10) only introduces one independent new condition. The symmetry conditions (7.9) and (7.10) therefore reduce the number of independent components of the Riemann tensor in four dimensions from $4^4 = 256$ to 20.

The Riemann tensor also satisfies the *Bianchi identities*

$$D_\lambda R^\rho{}_{\sigma\mu\nu} + D_\mu R^\rho{}_{\sigma\nu\lambda} + D_\nu R^\rho{}_{\sigma\lambda\mu} = 0, \tag{7.11}$$

which were helpful in the search for the correct dynamical equation for gravitational fields. Measures of local curvature which follow from the Riemann tensor are the *Ricci tensor*

$$R_{\mu\nu} = R^\rho{}_{\mu\rho\nu} = R_{\nu\mu}, \tag{7.12}$$

and the *curvature scalar* (or Ricci scalar)

$$R = R^\mu{}_\mu. \tag{7.13}$$

The divergence of the Ricci tensor can be written as a derivative of the curvature scalar: Contracting the Bianchi identity

$$D_\lambda R^\rho{}_{\mu\sigma\nu} + D_\sigma R^\rho{}_{\mu\nu\lambda} + D_\nu R^\rho{}_{\mu\lambda\sigma} = 0 \tag{7.14}$$

over ρ and σ yields

$$D_\lambda R_{\mu\nu} + D_\sigma R^\sigma{}_{\mu\nu\lambda} - D_\nu R_{\mu\lambda} = 0, \tag{7.15}$$

and contracting over λ and μ then yields

$$2D^\mu R_{\mu\nu} - \partial_\nu R = 0. \tag{7.16}$$

We will see in section 7.2 that this identity played a key role in finding the correct dynamical equation for the generation and evolution of gravitational fields.

In concluding this subsection, we note that the Riemann tensor also arises in the commutator of covariant derivatives,

$$\begin{aligned} [D_\mu, D_\nu]v^\rho &= \partial_\mu D_\nu v^\rho + \Gamma^\rho{}_{\sigma\mu} D_\nu v^\sigma - \Gamma^\sigma{}_{\nu\mu} D_\sigma v^\rho - (\mu \leftrightarrow \nu) \\ &= \left(\partial_\mu \Gamma^\rho{}_{\sigma\nu} - \partial_\nu \Gamma^\rho{}_{\sigma\mu} + \Gamma^\rho{}_{\lambda\mu}\Gamma^\lambda{}_{\sigma\nu} - \Gamma^\rho{}_{\lambda\nu}\Gamma^\lambda{}_{\sigma\mu}\right) v^\sigma \\ &= R^\rho{}_{\sigma\mu\nu} v^\sigma. \end{aligned} \quad (7.17)$$

7.2 The Einstein equation

We have seen that the geometry of spacetime determines the gravitational field through the Christoffel symbols, and the Christoffel symbols are determined by the metric. Since gravitational fields in Newtonian mechanics are dynamically generated through masses, the metric in general relativity must also be generated dynamically, and we would like to reveal the dynamical mechanisms which determine the geometry of spacetime. This means that we would like to identify a differential equation which connects the metric to its dynamical sources. This equation can be determined from the following observations:

- We know that in the limit of weak gravity the metric is determined by masses.
- We know from special relativity that mass is only a particular form of energy, and that energy is only a particular component of the 4-momentum of a physical system.
- Differential equations relate local quantities, i.e. the differential equation for the metric must relate the metric and its derivatives to the *densities* of energy and momentum.
- Densities of energy and momentum are components of the stress–energy tensor. The stress–energy tensor of the local energy and momentum distribution must therefore describe the sources of the local gravitational field.

Furthermore:

- The equation for the dynamical generation of the gravitational field should have the same form for each observer, i.e. it should be an equation between tensors.
- The only tensors which are linear in second-order derivatives of the metric are the Riemann tensor, the Ricci tensor, and the curvature scalar.
- The tensor equation should have ten components, since it should determine the ten independent components of $g_{\mu\nu}$. The equation must therefore somehow relate the Ricci tensor to the stress–energy tensor[1].

Furthermore:

[1] We cannot simply require $g_{\mu\nu} \propto T_{\mu\nu}$, because generically $D_\rho T_{\mu\nu} \neq 0$, and also because generically $T_\mu^\mu \equiv T \neq 4$.

- The stress–energy tensor satisfies $D^\mu T_{\mu\nu} = 0$. Therefore the curvature tensor related to $T_{\mu\nu}$ must satisfy this property as an identity, in order not to generate an extra higher-order equation for the metric.

The following combination of the Ricci tensor and the curvature scalar,

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R, \tag{7.18}$$

has the required property as a consequence of the Bianchi identities (7.11), cf equation (7.16),

$$D^\mu G_{\mu\nu} = D^\mu R_{\mu\nu} - \frac{1}{2}\partial_\nu R = 0. \tag{7.19}$$

$G_{\mu\nu}$ is the *Einstein tensor*, and the previous reasoning leads to the *Einstein equation*

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = \kappa T_{\mu\nu}. \tag{7.20}$$

This equation relates gravitational fields, i.e. the local geometry of spacetime, to its sources, i.e. the local distribution of energy–momentum densities and currents.

Taking the trace of equation (7.20) yields $R = -\kappa T$ and therefore the equivalent equation

$$R_{\mu\nu} = \kappa\left(T_{\mu\nu} - \frac{1}{2}g_{\mu\nu}T\right). \tag{7.21}$$

We can determine the proportionality constant κ in equations (7.20) and (7.21) through the weak field limit, which should yield Newtonian gravity:

The Newtonian limit of Einstein's equation—determination of κ

The weak field expansion $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$ of the metric implies in first order of $h_{\mu\nu}$ the equations

$$g^{\mu\nu} \simeq \eta^{\mu\nu} - \eta^{\mu\rho}\eta^{\nu\sigma}h_{\rho\sigma} = \eta^{\mu\nu} - h^{\mu\nu}, \tag{7.22}$$

$$\Gamma^\rho{}_{\nu\mu} \simeq \frac{1}{2}\eta^{\rho\sigma}\left(\partial_\nu h_{\sigma\mu} + \partial_\mu h_{\sigma\nu} - \partial_\sigma h_{\mu\nu}\right), \tag{7.23}$$

$$\Gamma^\rho{}_{\rho\mu} \simeq \frac{1}{2}\partial_\mu h^\rho{}_\rho,\ R^\rho{}_{\mu\sigma\nu} \simeq \partial_\sigma\Gamma^\rho{}_{\mu\nu} - \partial_\nu\Gamma^\rho{}_{\mu\sigma}, \tag{7.24}$$

and finally

$$\begin{aligned} R_{\mu\nu} &\simeq \partial_\rho\Gamma^\rho{}_{\mu\nu} - \partial_\nu\Gamma^\rho{}_{\mu\rho} \\ &\simeq \frac{1}{2}\left(\partial_\nu\partial^\rho h_{\rho\mu} + \partial_\mu\partial^\rho h_{\rho\nu} - \partial_\rho\partial^\rho h_{\mu\nu} - \partial_\mu\partial_\nu h^\rho{}_\rho\right). \end{aligned} \tag{7.25}$$

The linearized form of the Einstein equation (7.21) then reads

$$\partial_\rho\partial^\rho h_{\mu\nu} - \partial_\nu\partial^\rho h_{\rho\mu} - \partial_\mu\partial^\rho h_{\rho\nu} + \partial_\mu\partial_\nu h^\rho{}_\rho = -2\kappa\left(T_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}T\right). \tag{7.26}$$

We can simplify this equation by choosing coordinates in such a way that the left-hand side becomes simpler. In particular we can choose the coordinates such that the conditions of *harmonic gauge* (also called *de Donder gauge*) are fulfilled[2]:

$$\partial^\rho h_{\rho\mu} = \frac{1}{2}\partial_\mu h^\rho{}_\rho. \tag{7.27}$$

With these conditions equation (7.26) reduces to the four-dimensional wave equation

$$\partial_\rho\partial^\rho h_{\mu\nu} = -2\kappa\left(T_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}T\right). \tag{7.28}$$

The stress–energy tensor of a time-independent mass density ϱ_M is $T_{\mu\nu} = \varrho_M c^2\eta_\mu{}^0\eta_\nu{}^0$, $T = -\varrho_M c^2$, and substitution into the previous wave equation yields for $\mu = \nu = 0$ and for time-independent perturbation h_{00} the equation

$$\Delta h_{00} = -\kappa\varrho_M c^2. \tag{7.29}$$

We know already from the linearized geodesic equation that h_{00} is related to the Newtonian gravitational potential through $h_{00} = -(2/c^2)\Phi$, see equation (6.43), and therefore the previous equation reads

$$\Delta\Phi = \frac{1}{2}c^4\kappa\varrho_M. \tag{7.30}$$

Comparison with the Newtonian equation $\Delta\Phi = 4\pi G\varrho_M$ shows that

$$\kappa = \frac{8\pi G}{c^4} = \frac{\hbar}{m_{\text{Planck}}^2 c^3} = 2.076 \times 10^{-43}\frac{1}{\text{N}}. \tag{7.31}$$

Here m_{Planck} denotes the *reduced Planck mass*. The reduced Planck units are

$$m_{\text{Planck}} = \frac{M_{\text{Planck}}}{\sqrt{8\pi}} = \sqrt{\frac{\hbar c}{8\pi G}} = \frac{\hbar}{c\ell_{\text{Planck}}} = 2.435 \times 10^{18}\,\text{GeV}/c^2, \tag{7.32}$$

$$\begin{aligned}\ell_{\text{Planck}} = \sqrt{8\pi}\,L_{\text{Planck}} &= \frac{\hbar}{m_{\text{Planck}}c} = \sqrt{\frac{8\pi G\hbar}{c^3}} = \sqrt{\hbar c\kappa}\\ &= 8.103 \times 10^{-33}\text{cm} = ct_{\text{Planck}} = c \times 2.703 \times 10^{-43}\text{s}.\end{aligned} \tag{7.33}$$

[2] See the following subsection for the feasibility of this gauge.

Feasibility of the harmonic gauge

It is sufficient here to consider coordinate transformations $x^\mu \to x'^\mu = x^\mu - \epsilon^\mu$ in first order, since harmonic gauge itself pertains only to first-order perturbations of flat Minkowski space. The coordinate transformation $x'^\mu = x^\mu - \epsilon^\mu(x)$ yields $\partial'_\mu x^\alpha = \eta_\mu{}^\alpha + \partial_\mu \epsilon^\alpha$, and therefore the metric transforms according to

$$g'_{\mu\nu} = \partial'_\mu x^\alpha \, \partial'_\nu x^\beta \, g_{\alpha\beta} = g_{\mu\nu} + \partial_\mu \epsilon^\alpha \, g_{\alpha\nu} + \partial_\nu \epsilon^\alpha \, g_{\mu\alpha}. \tag{7.34}$$

This yields for the first-order terms $h_{\mu\nu}$

$$h'_{\mu\nu} = h_{\mu\nu} + \partial_\mu \epsilon_\nu + \partial_\nu \epsilon_\mu, \tag{7.35}$$

and

$$\partial^\sigma h'_{\sigma\nu} = \partial^\sigma h_{\sigma\nu} + \partial^\sigma \partial_\sigma \epsilon_\nu + \partial_\nu \partial^\sigma \epsilon_\sigma, \quad \partial_\nu h'^\sigma{}_\sigma = \partial_\nu h^\sigma{}_\sigma + 2\partial_\nu \partial^\sigma \epsilon_\sigma. \tag{7.36}$$

Equations (7.36) imply in particular

$$\partial^\sigma h'_{\sigma\nu} - \frac{1}{2}\partial_\nu h'^\sigma{}_\sigma = \partial^\sigma h_{\sigma\nu} - \frac{1}{2}\partial_\nu h^\sigma{}_\sigma + \partial^\sigma \partial_\sigma \epsilon_\nu. \tag{7.37}$$

Therefore, the harmonic gauge condition

$$\partial^\sigma h'_{\sigma\nu} - \frac{1}{2}\partial_\nu h'^\sigma{}_\sigma = 0 \tag{7.38}$$

in the new coordinates implies in the old coordinates the four-dimensional wave equations

$$\partial^\sigma \partial_\sigma \epsilon_\nu = \frac{1}{2}\partial_\nu h^\sigma{}_\sigma - \partial^\sigma h_{\sigma\nu}. \tag{7.39}$$

This can always be solved through a Green's function,

$$\epsilon_\nu(x) = \int d^4x' \, G(x - x')\left(\partial^\sigma h_{\sigma\nu}(x') - \frac{1}{2}\partial_\nu h^\sigma{}_\sigma(x')\right), \tag{7.40}$$

$$G(x) = \frac{1}{4\pi r}\delta(r - ct), \quad \partial^\sigma \partial_\sigma G(x) = -\delta(x). \tag{7.41}$$

The Einstein equation with a cosmological constant

There is another geometric tensor besides the Einstein tensor which contains at most second-order derivatives of the metric and satisfies the constraint of vanishing covariant divergence: The covariant constancy of the metric, $D_\rho g_{\mu\nu} = 0$, implies $D^\mu g_{\mu\nu} = 0$. Therefore we can include a term proportional to the metric in the Einstein equation,

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R + \lambda g_{\mu\nu} = \kappa T_{\mu\nu}, \tag{7.42}$$

or

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = \kappa T_{\mu\nu} - \kappa\Lambda g_{\mu\nu}, \tag{7.43}$$

with a constant $\lambda = \kappa\Lambda$. λ (or equivalently Λ) is denoted as a *cosmological constant*.

Including the cosmological constant on the left-hand side (λ) of the Einstein equation portrays it as part of the geometry—as a permissible shift of the Einstein tensor. With the definitions

$$G^{(0)}_{\mu\nu} = -\lambda g_{\mu\nu}, \quad R^{(0)}_{\mu\nu} = \lambda g_{\mu\nu}, \tag{7.44}$$

one can think of the Einstein equation in the version (7.42) as a perturbation of the background spacetime geometry due to the matter energy–momentum tensor,

$$G_{\mu\nu} - G^{(0)}_{\mu\nu} = R_{\mu\nu} - R^{(0)}_{\mu\nu} - \frac{1}{2}g_{\mu\nu}(R - R^{(0)}) = \kappa T_{\mu\nu}. \tag{7.45}$$

A background spacetime of the form (7.44) is denoted as an *Einstein space*.

Including the cosmological constant in the form (7.43) portrays it as part of the energy–momentum tensor

$$T^{(\Lambda)}_{\mu\nu} = -\Lambda g_{\mu\nu} \tag{7.46}$$

corresponding to a constant energy density $\varrho = \Lambda$ and constant pressure $p = -\Lambda = -\varrho$.

The two points of view are mathematically completely equivalent, of course, but offer interesting differences in perspective.

7.3 The Schwarzschild metric: The gravitational field outside a non-rotating star

In 1916 Karl Schwarzschild[3] calculated the metric outside a non-rotating spherical mass distribution of mass M,

$$ds^2 = -\left(1 - \frac{2GM}{c^2 r}\right)c^2 dt^2 + \frac{1}{1 - \frac{2GM}{c^2 r}}dr^2 + r^2 d\vartheta^2 + r^2 \sin^2\vartheta\, d\varphi^2. \tag{7.47}$$

The mass dependent length scale

$$r_S = 2GM/c^2 \tag{7.48}$$

is denoted as the *Schwarzschild radius* of the mass M.

The angular part of the metric tells us that the length of a circle with radial coordinate r still $2\pi r$, and the area of the sphere with radial coordinate r is $4\pi r^2$.

However, the radial coordinate r is not a radial distance, and even becomes a timelike coordinate for $r < r_S$. The radial distance element for $r > r_S$ is

[3] Schwarzschild K 1916 *Sitzungsberichte der Königlich-Preussischen Akademie der Wissenschaften* 189.

$d\ell = \sqrt{r/(r - r_S)}\,dr > dr$, i.e. the distance between spherical shells of areas $4\pi r_0^2$ and $4\pi r_1^2$ is larger than in the Minkowski spacetime, or stated differently, angular size grows more slowly with distance in the Schwarzschild spacetime. Radial distances are stretched near the Schwarzschild radius $r \gtrsim r_S$. Indeed, the radial distance between points with radial coordinates $r_1 \geqslant r_0 \geqslant r_S$ is

$$\begin{aligned}\ell(r_1, r_0) &= \int_{r_0}^{r_1} dr\sqrt{\frac{r}{r - r_S}} = \left[\sqrt{r(r - r_S)} + r_S \ln\left(\sqrt{r} + \sqrt{r - r_S}\right)\right]_{r_0}^{r_1} \\ &= \sqrt{r_1(r_1 - r_S)} - \sqrt{r_0(r_0 - r_S)} + r_S \ln\left(\frac{\sqrt{r_1} + \sqrt{r_1 - r_S}}{\sqrt{r_0} + \sqrt{r_0 - r_S}}\right).\end{aligned} \tag{7.49}$$

This distance exceeds $r_1 - r_0$, but approaches $r_1 - r_0$ for $r_1 \geqslant r_0 \gg r_S$ in the sense $\lim_{(r_1/r_S) > (r_0/r_S) \to \infty} \ell(r_1, r_0)/(r_1 - r_0) \to 1$.

To visualize the effects of curvature, we can take two-dimensional sections of spacetime and embed them in an auxiliary three-dimensional flat space. To visualize the effect on the radial distances in the Schwarzschild spacetime, we can take the $\{x, y\}$ plane and embed it in an auxiliary flat $\{x, y, Z\}$ space, $\{x, y\} \to Z(x, y)$, such that the embedded surface has distances between radial coordinates r_0 and r_1 which correspond to the radial distances *on* the $\{x, y\}$ plane in the Schwarzschild spacetime in the region $r = \sqrt{x^2 + y^2} > r_S$:

$$d\ell^2 = \frac{r}{r - r_S} dr^2 = \left[1 + \left(\frac{dZ}{dr}\right)^2\right] dr^2. \tag{7.50}$$

This yields $Z(r) = 2\sqrt{r_S(r - r_S)}$ and corresponds to the funnel surface displayed in figure 7.1.

The disc $r = \sqrt{x^2 + y^2} < r_S$ corresponds to timelike radial coordinate and can be embedded into a three-dimensional Minkowski spacetime with spacelike coordinates $\{x, y\}$ and a timelike coordinate Z,

$$-c^2 dT^2 = \frac{r}{r - r_S} dr^2 = \left[1 - \left(\frac{dZ}{dr}\right)^2\right] dr^2. \tag{7.51}$$

This can be solved by $Z(r) = -2\sqrt{r_S(r_S - r)}$ and yields the timelike cone displayed in figure 7.1. However, note that the singular point $r = 0$, $Z = -2r_S$ at the tip of the timelike cone still corresponds to an infinitely extended spatial line $-\infty < t < \infty$, with diverging length element $dL = cdt\sqrt{(r_S - r)/r}$.

The upper part $Z > 0$ in figure 7.1 can be used to visualize spatial distances in the orbital plane for $r \geqslant r_S$: Distances between any two points on the funnel surface are the same distances as on the orbital plane in the Schwarzschild spacetime. However, one has to exercise caution when using this picture for particle motion in the Schwarzschild spacetime, because the orbits of particles in the Schwarzschild equatorial plane are geodesics of the three-dimensional metric

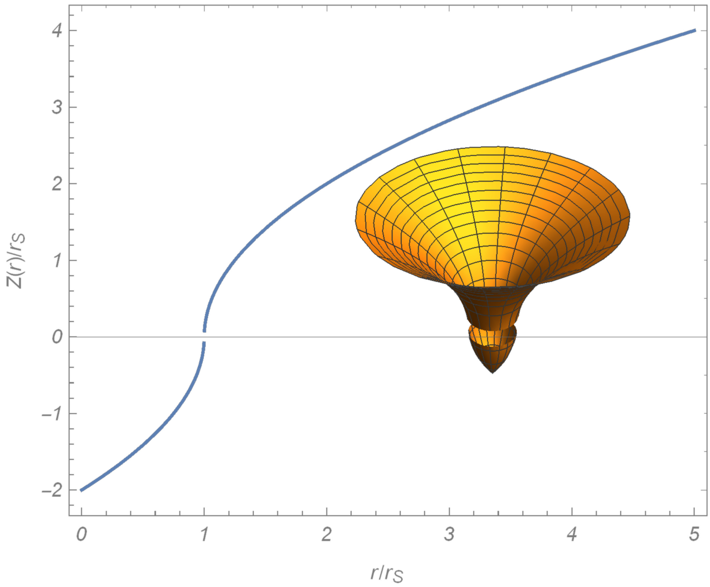

Figure 7.1. The embeddings of regions of the $\{x, y\}$ plane of the Schwarzschild spacetime in an auxiliary three-dimensional flat space ($Z > 0$) and in an auxiliary three-dimensional Minkowski spacetime ($Z < 0$).

$$ds^2 = -\frac{r - r_S}{r}c^2dt^2 + d\ell^2(r) + r^2d\varphi^2, \tag{7.52}$$

they are *not* geodesics of the two-dimensional metric $d\ell^2(r) + r^2d\varphi^2$.

Furthermore, note that the regions near and below r_S do not exist in typical realizations of the Schwarzschild metric outside non-rotating planets or stars, since the radii of these objects are much larger than their Schwarzschild radii and the Schwarzschild metric does not describe the spacetime inside of a massive object. If the Schwarzschild horizon is outside of the mass M, the Schwarzschild metric describes a *black hole* for reasons which will be discussed later.

We will derive the Schwarzschild metric from the Einstein equation and then discuss its physical implications.

The spherically symmetric vacuum solution of the Einstein equation

This subsection presents a slightly modified and extended version of Weinberg's excellent derivation of the Schwarzschild metric[4].

As a definition of a spherically symmetric metric, we require the existence of coordinates t, r, ϑ, φ such that the surfaces $t = \text{const.}$, $r = \text{const.}$ are spheres of area $4\pi r^2$:

$$ds^2|_{dt=0,\, dr=0} = r^2(d\vartheta^2 + \sin^2\vartheta\, d\varphi^2), \tag{7.53}$$

and we also require that the full metric ds^2 is invariant under rotations.

[4] Weinberg S 1972 *Gravitation and Cosmology* (New York: Wiley).

The latter requirement excludes terms proportional to $dtd\vartheta$ and $dtd\varphi$ in ds^2, and implies that the remaining components of the metric can only depend on t and r:

$$g_{tt} = -B(r, t), \quad g_{tr} = C(r, t), \quad g_{rr} = A(r, t), \tag{7.54}$$

$$ds^2 = -B(r, t)c^2dt^2 + 2C(r, t)dtdr + A(r, t)dr^2 + r^2(d\vartheta^2 + \sin^2\vartheta\, d\varphi^2). \tag{7.55}$$

However, the function $C(t, r)$ can always be removed by a transformation of the coordinate t,

$$t \to t'(t, r): \quad dt' \propto dt - \frac{C(r, t)}{c^2B(r, t)}dr, \tag{7.56}$$

because there always exists an integrating factor $\alpha(r, t)$ such that

$$dt' = \alpha(r, t)\left(dt - \frac{C(r, t)}{c^2B(r, t)}dr\right), \tag{7.57}$$

$$\frac{\partial}{\partial r}\alpha(r, t) = -\frac{\partial}{\partial t}\left(\alpha(r, t)\frac{C(r, t)}{c^2B(r, t)}\right), \tag{7.58}$$

$$-Bc^2dt^2 + 2Cdtdr = -\frac{B}{\alpha^2}c^2dt'^2 + \frac{C^2}{c^2B}dr^2. \tag{7.59}$$

This leaves us with the following most general possibility of a spherically symmetric metric:

$$ds^2 = -B(r, t)c^2dt^2 + A(r, t)dr^2 + r^2(d\vartheta^2 + \sin^2\vartheta\, d\varphi^2), \tag{7.60}$$

with the following non-vanishing Christoffel symbols[5] (modulo symmetry in the lower indices),

$$\Gamma^r{}_{rr} = \frac{A'}{2A}, \quad \Gamma^r{}_{\vartheta\vartheta} = -\frac{r}{A}, \quad \Gamma^r{}_{\varphi\varphi} = -\frac{r\sin^2\vartheta}{A}, \tag{7.62}$$

$$\Gamma^\vartheta{}_{\vartheta r} = \Gamma^\varphi{}_{\varphi r} = \frac{1}{r}, \quad \Gamma^\vartheta{}_{\varphi\varphi} = -\sin\vartheta\cos\vartheta, \quad \Gamma^\varphi{}_{\varphi\vartheta} = \cot\vartheta, \tag{7.63}$$

$$\Gamma^r{}_{tt} = \frac{c^2B'}{2A}, \quad \Gamma^t{}_{tr} = \frac{B'}{2B}, \quad \Gamma^t{}_{tt} = \frac{\dot{B}}{2B}, \quad \Gamma^r{}_{rt} = \frac{\dot{A}}{2A}, \tag{7.64}$$

[5] Recall that the Christoffel symbols can be calculated from the metric using the last line in equation (5.13). However, a faster method is to derive the geodesic equation from the eigentime Lagrange function (6.23), which here takes the form

$$2L_\tau/m = -B(r, t)(dt/d\tau)^2 + A(r, t)(dr/d\tau)^2 + r^2(d\vartheta/d\tau)^2 + r^2\sin^2\vartheta(d\varphi/d\tau)^2. \tag{7.61}$$

$$\Gamma^t{}_{rr} = \frac{\dot{A}}{2c^2 B}. \tag{7.65}$$

The Christoffel symbols yield in particular the following component of the Ricci tensor:

$$\begin{aligned} R_{rt} &= \partial_r \Gamma^r{}_{rt} + \partial_t \Gamma^t{}_{rt} - \partial_t\left(\Gamma^r{}_{rr} + \Gamma^t{}_{tr} + \Gamma^\vartheta{}_{\vartheta r} + \Gamma^\varphi{}_{\varphi r}\right) + \left(\Gamma^t{}_{tt} + \Gamma^r{}_{rt}\right)\Gamma^t{}_{rt} \\ &\quad + \left(\Gamma^r{}_{rr} + \Gamma^\vartheta{}_{\vartheta r} + \Gamma^\varphi{}_{\varphi r} + \Gamma^t{}_{tr}\right)\Gamma^r{}_{rt} - \Gamma^r{}_{rr}\Gamma^r{}_{rt} - \Gamma^r{}_{tr}\Gamma^t{}_{rt} - \Gamma^t{}_{rr}\Gamma^r{}_{tt} \\ &\quad - \Gamma^t{}_{tr}\Gamma^t{}_{tt} = \frac{\dot{A}}{Ar}. \end{aligned} \tag{7.66}$$

The Einstein equations outside of the star are $R_{\mu\nu} = 0$, and the equation $R_{rt} = 0$ then implies $\dot{A} = 0$, which also yields $\Gamma^r{}_{rt} = 0$, $\Gamma^t{}_{rr} = 0$.

The remaining non-vanishing components of the Ricci tensor are then

$$\frac{1}{c^2} R_{tt} = \frac{B''}{2A} - \frac{A'B'}{4A^2} + \frac{B'}{Ar} - \frac{B'^2}{4AB}, \quad R_{rr} = -\frac{B''}{2B} + \frac{B'^2}{4B^2} + \frac{A'B'}{4AB} + \frac{A'}{Ar}, \tag{7.67}$$

$$R_{\vartheta\vartheta} = 1 - \frac{1}{A} + \frac{rA'}{2A^2} - \frac{rB'}{2AB}, \qquad R_{\varphi\varphi} = \sin^2\vartheta\, R_{\vartheta\vartheta}. \tag{7.68}$$

The equation

$$\frac{1}{c^2 B} R_{tt} + \frac{1}{A} R_{rr} = \frac{1}{Ar}\left(\frac{B'}{B} + \frac{A'}{A}\right) = \frac{1}{A^2 Br}(AB)' = 0 \tag{7.69}$$

yields

$$B(r, t) = \frac{f(t)}{A(r)}. \tag{7.70}$$

However, we can gauge $f(t)$ away through a redefinition of t: $dt \to dt' = \sqrt{f(t)}\,dt$, and we end up with a completely t-independent metric:

$$B(r) = \frac{1}{A(r)}. \tag{7.71}$$

The result that the gravitational field outside of a spherically symmetric (but possibly t-dependent) source does not depend on t (and is therefore static in the region $r > r_S$) is the *(Jebsen–)Birkhoff theorem*[6]. Substitution of equation (7.71) into $R_{\vartheta\vartheta}$ yields

[6] Jebsen J T 1921 *Arkiv för Matematik, Astronomi och Fysik* **15** 18; Birkhoff G D and Langer R E 1923 *Relativity and Modern Physics* (Cambridge, MA: Harvard University Press); Schmidt H-J 1997 *Grav. Cosmol.* **3** 185; Deser S 2005 *Gen. Relat. Grav.* 37 2251.

$$R_{\vartheta\vartheta} = 1 - \frac{1}{A} + \frac{rA'}{A^2} = 1 - \frac{d}{dr}\frac{r}{A} = 0, \tag{7.72}$$

which implies

$$A = \frac{r}{r+\zeta}, \quad B = 1 + \frac{\zeta}{r}, \tag{7.73}$$

where ζ is an integration constant. We know from our study of the Newtonian limit of the geodesic equation that $-g_{00} - 1 = B - 1$ must approach the Newtonian potential of the mass M,

$$B(r)|_{r\to\infty} \to 1 + \frac{2}{c^2}\Phi = 1 - \frac{2GM}{c^2 r}, \tag{7.74}$$

and this determines the integration constant ζ in terms of M,

$$\zeta = -\frac{2GM}{c^2} = -r_S. \tag{7.75}$$

We have found the *Schwarzschild metric* as the *unique spherically symmetric vacuum solution of the Einstein equation*:

$$ds^2 = -\left(1 - \frac{r_S}{r}\right)c^2dt^2 + \frac{r}{r - r_S}dr^2 + r^2(d\vartheta^2 + \sin^2\vartheta\, d\varphi^2). \tag{7.76}$$

The non-vanishing Christoffel symbols (up to symmetry in the lower indices) are

$$\Gamma^r{}_{rr} = -\frac{r_S}{2r(r-r_S)}, \quad \Gamma^r{}_{\vartheta\vartheta} = -(r - r_S), \tag{7.77}$$

$$\Gamma^r{}_{\varphi\varphi} = -(r - r_S)\sin^2\vartheta, \quad \Gamma^r{}_{tt} = c^2 r_S\frac{r - r_S}{2r^3}, \quad \Gamma^\vartheta{}_{\vartheta r} = \Gamma^\varphi{}_{\varphi r} = \frac{1}{r}, \tag{7.78}$$

$$\Gamma^\vartheta{}_{\varphi\varphi} = -\sin\vartheta\cos\vartheta, \quad \Gamma^\varphi{}_{\varphi\vartheta} = \cot\theta, \quad \Gamma^t{}_{tr} = \frac{r_S}{2r(r-r_S)}. \tag{7.79}$$

The non-vanishing components of the Riemann tensor are up to the symmetries (7.9) given by

$$R_{r\vartheta r\vartheta} = -\frac{r_S}{2(r-r_S)}, \quad R_{r\varphi r\varphi} = -\frac{r_S}{2(r-r_S)}\sin^2\vartheta, \tag{7.80}$$

$$R_{rtrt} = -c^2\frac{r_S}{r^3}, \quad R_{\vartheta t\vartheta t} = c^2 r_S\frac{r - r_S}{2r^2}, \quad R_{\varphi t\varphi t} = c^2 r_S\frac{r - r_S}{2r^2}\sin^2\vartheta, \tag{7.81}$$

$$R_{\vartheta\varphi\vartheta\varphi} = r r_S \sin^2\vartheta. \tag{7.82}$$

Some of the metric coefficients in equation (7.76), some of the resulting Christoffel symbols, and some of the resulting curvature components are singular at $r = 0$ or at $r = r_S$. However, the singularity at $r = r_S$ can be removed through coordinate transformations. We will see an explicit example of this in section 7.4. On the other hand, the singularity at $r = 0$ cannot be removed by any coordinate transformation. We can recognize this by calculating the coordinate-independent curvature measure

$$R^{\mu\nu\rho\sigma}R_{\mu\nu\rho\sigma} = 12r_S^2/r^6. \tag{7.83}$$

Some curvature components (and therefore also tidal forces) will always diverge for $r \to 0$.

7.4 The interior of Schwarzschild black holes

Besides having singular metric coefficients at $r = r_S$, the coordinates r and t also exchange their roles for $r < r_S$: The sign change in the metric components implies that t is a spatial coordinate and $-r$ is a timelike coordinate inside the horizon.

In 1960 Martin Kruskal[7] and George Szekeres[8] independently solved the following problem: Find a coordinate transformation $r, t \Rightarrow u, v$ such that

1. the metric at the event horizon is non-singular in the new coordinates,
2. v should remain timelike and u should remain spacelike everywhere, and
3. the world lines of radially moving photons ($d\varphi = d\vartheta = 0$) should satisfy $dv = \pm du$.
4. **Inside** the horizon we therefore require[9]

$$\frac{r_S - r}{r}dt^2 - \frac{r}{r_S - r}dr^2 = C(u, v)(du^2 - dv^2), \tag{7.84}$$

with a positive definite proportionality factor $C(u, v)$, which may diverge only for $r = 0$. This implies for the partial derivatives of $u(t, r)$ and $v(t, r)$

$$\frac{\partial u}{\partial r}\frac{\partial u}{\partial t} = \frac{\partial v}{\partial r}\frac{\partial v}{\partial t}, \tag{7.85}$$

$$\left(\frac{\partial u}{\partial t}\right)^2 - \left(\frac{\partial v}{\partial t}\right)^2 = \frac{1}{C}\frac{r_S - r}{r}, \quad \left(\frac{\partial u}{\partial r}\right)^2 - \left(\frac{\partial v}{\partial r}\right)^2 = -\frac{1}{C}\frac{r}{r_S - r}. \tag{7.86}$$

The last two equations imply

$$\begin{aligned}\left(\frac{\partial u}{\partial t}\right)^2 - \left(\frac{\partial v}{\partial t}\right)^2 &= -\left(\frac{r_S - r}{r}\right)^2\left[\left(\frac{\partial u}{\partial r}\right)^2 - \left(\frac{\partial v}{\partial r}\right)^2\right] \\ &= -\left[\left(\frac{\partial u}{\partial \xi}\right)^2 - \left(\frac{\partial v}{\partial \xi}\right)^2\right],\end{aligned} \tag{7.87}$$

with

[7] Kruskal M D 1960 *Phys. Rev.* 119 1743.
[8] Szekeres G 1960 *Pub. Math. Debrecen* **7** 285.
[9] Here we use $c = 1$.

$$\frac{\partial}{\partial\xi} = \frac{r_S - r}{r}\frac{\partial}{\partial r} \quad \Rightarrow \quad \xi = -r_S \ln\frac{r_S - r}{r_S} - r. \tag{7.88}$$

The equations (7.85) and (7.87) are equivalent to

$$\left(\frac{\partial u}{\partial t} \pm \frac{\partial u}{\partial \xi}\right)^2 = \left(\frac{\partial v}{\partial t} \pm \frac{\partial v}{\partial \xi}\right)^2. \tag{7.89}$$

Since we must have $u \neq \pm v$ we cannot take the same sign for the square roots in both equations, and there remains only one possibility (up to a relative sign change between u and v):

$$\frac{\partial u}{\partial t} + \frac{\partial u}{\partial \xi} = -\left(\frac{\partial v}{\partial t} + \frac{\partial v}{\partial \xi}\right), \quad \frac{\partial u}{\partial t} - \frac{\partial u}{\partial \xi} = \left(\frac{\partial v}{\partial t} - \frac{\partial v}{\partial \xi}\right), \tag{7.90}$$

corresponding to

$$u(t, \xi) = f(t + \xi) + g(t - \xi), \quad v(t, \xi) = -f(t + \xi) + g(t - \xi), \tag{7.91}$$

and a constant of proportionality in equation (7.84):

$$C(u, v) = \frac{r_S - r}{r(\dot{u}^2 - \dot{v}^2)} = \frac{r_S - r}{4r\dot{f}\,\dot{g}}. \tag{7.92}$$

We want the metric to be non-singular in $r = r_S$, i.e. we want to have

$$\dot{f}(t + \xi)\dot{g}(t - \xi) \propto r_S - r. \tag{7.93}$$

However, due to

$$\xi + r_S = -r_S \ln\frac{r_S - r}{r_S} + r_S - r \tag{7.94}$$

$r_S - r$ is a function of $2\xi = t + \xi - (t - \xi)$. We get an additive function in $\xi + t$ and $\xi - t$ from the product (7.93) of a function of $\xi + t$ and a function of $\xi - t$ by taking exponential functions of the arguments, and therefore we choose

$$f(t + \xi) = -\frac{1}{2}\exp\left(-\frac{t + \xi}{2r_S}\right) = -\frac{1}{2}\sqrt{\frac{r_S - r}{r_S}}\exp\left(-\frac{t - r}{2r_S}\right) \tag{7.95}$$

and

$$g(t - \xi) = \frac{1}{2}\exp\left(\frac{t - \xi}{r_S}\right) = \frac{1}{2}\sqrt{\frac{r_S - r}{r_S}}\exp\left(\frac{t + r}{2r_S}\right). \tag{7.96}$$

This yields

$$u(t, r) = \sqrt{\frac{r_S - r}{r_S}}\exp\left(\frac{r}{2r_S}\right)\sinh\left(\frac{t}{2r_S}\right) \tag{7.97}$$

and

$$v(t, r) = \sqrt{\frac{r_S - r}{r_S}}\exp\left(\frac{r}{2r_S}\right)\cosh\left(\frac{t}{2r_S}\right). \tag{7.98}$$

These are the *Kruskal–Szekeres coordinates* for the region $r < r_S$. A similar calculation in the region $r > r_S$ requiring

$$\frac{r}{r - r_S}dr^2 - \frac{r - r_S}{r}dt^2 = C(u, v)(du^2 - dv^2) \tag{7.99}$$

yields the Kruskal–Szekeres coordinates **outside the horizon**,

$$u(t, r) = \sqrt{\frac{r - r_S}{r_S}} \exp\left(\frac{r}{2r_S}\right)\cosh\left(\frac{t}{2r_S}\right), \tag{7.100}$$

$$v(t, r) = \sqrt{\frac{r - r_S}{r_S}} \exp\left(\frac{r}{2r_S}\right)\sinh\left(\frac{t}{2r_S}\right). \tag{7.101}$$

The metric in the Kruskal–Szekeres coordinates is

$$ds^2 = \frac{4r_S^3}{r} \exp\left(-\frac{r}{r_S}\right)(du^2 - dv^2) + r^2 d\vartheta^2 + r^2 \sin^2\vartheta\, d\varphi^2, \tag{7.102}$$

where $r = r(u, v)$ is the solution of

$$u^2 - v^2 = \frac{r - r_S}{r_S} \exp\left(\frac{r}{r_S}\right). \tag{7.103}$$

The (r, t) sections of the Schwarzschild spacetime in the (u, v) plane

Equation (7.103) shows that lines of constant r are hyperbolas in the (u, v) plane.

Outside the horizon we have $u > |v|$ according to the equations (7.100) and (7.101), and therefore the hyperbolas $r > r_S$ are single hyperbolic branches opening to the right in the positive u direction, see figure 7.2. The intersection of these hyperbolas with the u-axis is at

$$u = \sqrt{\frac{r - r_S}{r_S}} \exp\left(\frac{r}{2r_S}\right), \tag{7.104}$$

and the limiting hyperbola for $r \to r_S$ corresponds to the two half-lines which satisfy $u = |v|$.

Since $r = \text{const.}$ on the hyperbolas (7.103), these are t-coordinate lines with t increasing upwards along the hyperbolas, except for the limiting hyperbolas for $r \to r_S$: The half-line $u = -v > 0$ corresponds to $t \to \infty$ and the half-line $u = v > 0$ corresponds to $t \to \infty$.

Inside the horizon we have $v > |u|$ according to equations (7.97) and (7.98), and therefore the hyperbolas $r < r_S$ are single hyperbolic branches opening in positive v direction, see figure 7.3. The intersection of these hyperbolas with the v-axis is at

$$v = \sqrt{\frac{r_S - r}{r_S}} \exp\left(\frac{r}{2r_S}\right), \tag{7.105}$$

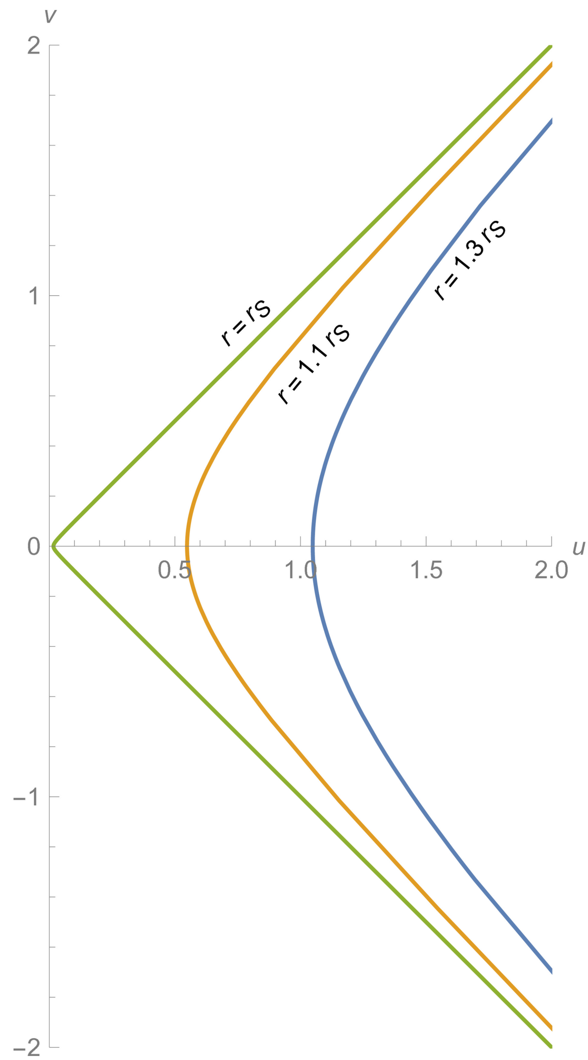

Figure 7.2. t-coordinate lines (r = const.) in the region $r > r_S$.

The singularity $r = 0$ corresponds to a spacelike hyperbola intersecting the v-axis in $v = 1$. The limiting hyperbola for $r \to r_S$ corresponds to the two half-lines $v > 0$, $u = -v$ ($t \to -\infty$ and $v > 0$, $u = v$ ($t \to \infty$).

The equations (7.100) and (7.101) yield for the t-coordinate outside the horizon

$$\tanh\left(\frac{t}{2r_S}\right) = \frac{v}{u}, \tag{7.106}$$

and therefore lines of constant t are straight half-lines in the quadrant $u > |v|$ with slopes

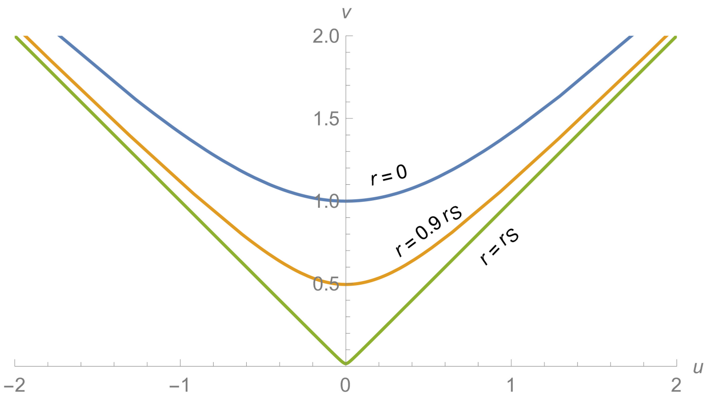

Figure 7.3. *t*-coordinate lines (r = const.) in the region $r < r_S$. The spatial coordinate t increases to the right along the hyperbolas, except for the singular hyperbola $r = r_S$, where $t = -\infty$ ($u < 0$) or $t = \infty$ ($u > 0$).

$$-1 < \frac{v}{u} = \tanh\left(\frac{t}{2r_S}\right) < 1. \tag{7.107}$$

On the other hand, the equations (7.97) and (7.98) yield for the t-coordinate inside the horizon

$$\tanh\left(\frac{t}{2r_S}\right) = \frac{u}{v}, \tag{7.108}$$

and therefore the lines of constant t inside the horizon are straight half-lines in the quadrant $v > |u|$ with inverse slopes

$$-1 < \frac{u}{v} = \tanh\left(\frac{t}{2r_S}\right) < 1. \tag{7.109}$$

The r-coordinate lines t = const. are depicted in figure 7.4.

Since r decreases with increasing v in the quadrant $v > |u|$, we infer that $-r$ is the timelike Schwarzschild coordinate inside the horizon.

From our outside observer interpretation of the radial coordinate we might have expected that the singularity $r = 0$ corresponds to the singular timelike world line traced out by a singular point in spacetime. However, figures 7.3 and 7.4 illustrate that the singularity corresponds to a spacelike line of singular moments, see also the discussion following equation (7.111).

The full Schwarzschild spacetime then corresponds to the region in the (u, v) plane which is bounded by the line $t = -\infty$ and the hyperbola $r = 0$, i.e. the region is determined by the conditions $u + v \geqslant 0$ and $v \leqslant \sqrt{1 + u^2}$.

The actual Schwarzschild horizon of the black hole corresponds to the line $r = r_S$, $t = \infty$, because this line separates the exterior asymptotically flat Universe from the singularity at $r = 0$. The fact that this line is degenerate in terms of Schwarzschild (r, t) coordinates is an expression of the singularity of Schwarzschild coordinates on the horizon. In the Kruskal–Szekeres coordinates each point on the line $r = r_S$, $t = \infty$ corresponds to the sphere $r \to r_S$ at different times v, at least from

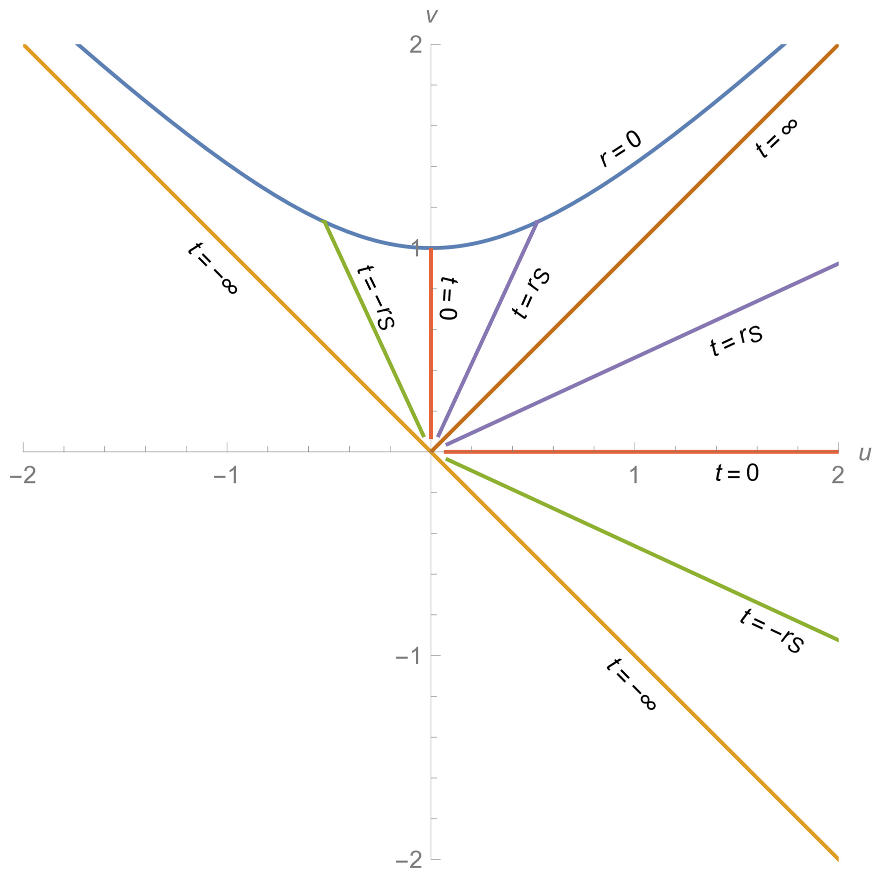

Figure 7.4. r-coordinate lines. The coordinate r increases with increasing u along the half-lines t = const. in the quadrant $u > |v|$ (i.e. $r > r_S$). The coordinate r decreases with increasing v along the half-lines $t = const.$ in the quadrant $v > |u|$ (i.e. $r < r_S$). The half-line $t = \infty$ and the line $t = -\infty$ have $r = r_S$.

our outside point of view. From the inside point of view the line $r = r_S$, $t = \infty$ constitutes one half of the limiting spacelike hypersurface for time parameter $-r \to -r_S$. Equation (7.102) shows that distances on this hypersurface are given by

$$d\ell^2 = r_S^2(d\vartheta^2 + \sin^2\vartheta\, d\varphi^2), \tag{7.110}$$

and therefore this limiting spacelike hypersurface still has a spherical component. However, distances on the spacelike hypersurfaces at later times $-r = \boldsymbol{const.}$ with $-\mathrm{r}_S < -\mathrm{r} < 0$ are given by

$$d\ell^2 = \frac{r_S - r}{r}dt^2 + r^2(d\vartheta^2 + \sin^2\vartheta\, d\varphi^2). \tag{7.111}$$

This metric corresponds to the topology of a line times a sphere $\mathbb{R}_1 \times S_2$, which is a three-dimensional infinitely long cylinder hypersurface (of a four-dimensional bulk cylinder, but the radial dimension of the cylinder in this case is timelike). The hypersurface $r = r_S$ is also a three-dimensional cylinder hypersurface, but with lightlike axis. Inside the horizon we start at time $-r = -r_S$ with the cylinder with lightlike axis and two-dimensional spacelike spherical surface sections $r_S^2(d\vartheta^2 + \sin^2\vartheta\, d\varphi^2)$, and go through three-dimensional spacelike cylinder hypersurfaces $\mathbb{R}_1 \times S_2$ with shrinking two-dimensional surface sections S_2 to a singular spacelike line with diverging metric $g_{tt} \to \infty$. This means for an infalling observer, inside the Schwarzschild horizon they will intersect three-dimensional cylinder hypersurfaces $-r = \text{const.}$ (7.111) for smaller and smaller values of r (increasing $-r$), which implies that the spherical surface components of the cylinder hypersurfaces shrink during their journey. Finally they will hit the singularity $r = 0$ at some finite value of the time v. At this point they will be infinitely stretched in the spatial t-direction, while their extensions in the (ϑ, φ)-directions are completely squeezed (i.e. they have been 'spaghettified'). You can more easily imagine this in terms of a two-dimensional cylinder surface

$$d\ell^2 = \frac{r_S - r}{r}dt^2 + r^2 d\varphi^2, \tag{7.112}$$

which is the section $\vartheta = \pi/2$ of the hypersurface (7.111). The spatial dimensions of the infalling astronaut's body would cover intervals Δt and $\Delta\varphi$, and approaching the singularity is like going into an infinitely long pinch at the end $r = 0$ of the cylinder surface (7.112).

7.5 Maximal extension of the Schwarzschild spacetime and wormholes

The Schwarzschild spacetime described in the previous picture is 'geodesically incomplete', which means that it contains geodesics which start at the boundary $t = -\infty$ with finite value of the corresponding eigentime or affine parameter τ, respectively.

A spacetime where all geodesics either extend to infinity or stop or emerge in a singularity is denoted as geodesically complete, and the Schwarzschild spacetime can be extended to a complete spacetime by adding the region bounded by $t = -\infty$, $r = r_S$ and the lower parabola $r = 0$ in figure 7.5.

Here corresponding Schwarzschild coordinates can be introduced in the new regions through:

$u < -|v|$ (new region outside horizons, $r > r_S$):

$$u(t, r) = -\sqrt{\frac{r - r_S}{r_S}}\exp\left(\frac{r}{2r_S}\right)\cosh\left(\frac{t}{2r_S}\right), \tag{7.113}$$

$$v(t, r) = -\sqrt{\frac{r - r_S}{r_S}}\exp\left(\frac{r}{2r_S}\right)\sinh\left(\frac{t}{2r_S}\right). \tag{7.114}$$

$-\sqrt{1 + u^2} \leqslant v < -|u|$ (new region inside horizons around the second singularity $r = 0$):

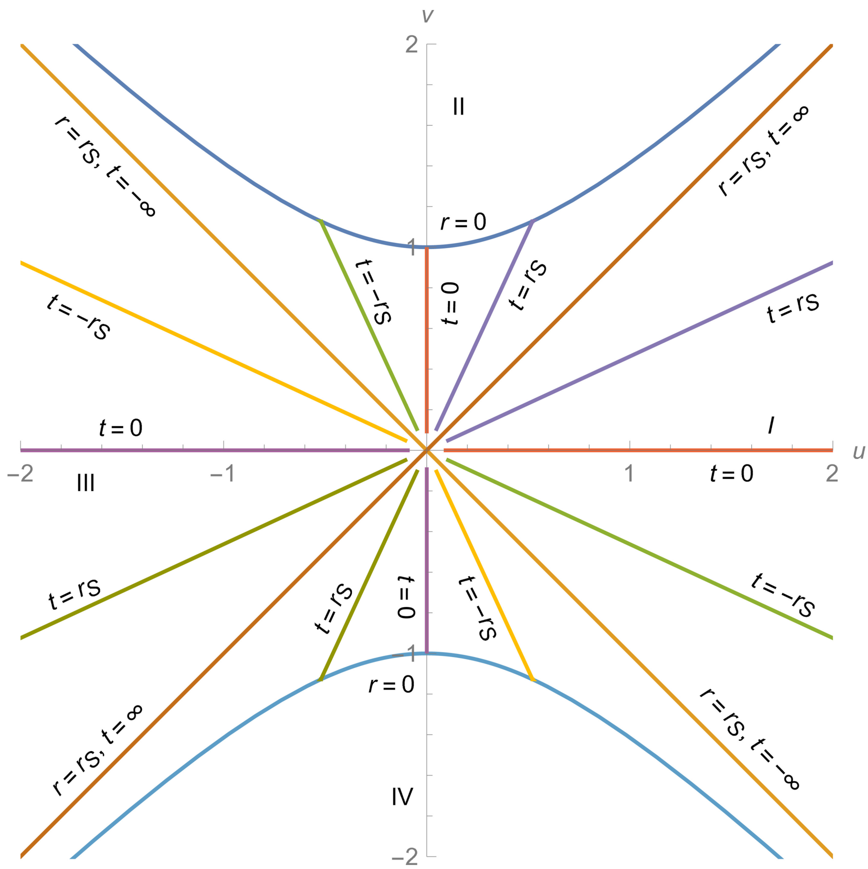

Figure 7.5. r-coordinate lines in a geodesic completion of the Schwarzschild spacetime. Starting at $r = 0$, r increases along lines of identical color, except for the horizons $r = r_S$, $t = \pm\infty$.

$$u(t, r) = -\sqrt{\frac{r_S - r}{r_S}}\exp\left(\frac{r}{2r_S}\right)\sinh\left(\frac{t}{2r_S}\right), \tag{7.115}$$

$$v(t, r) = -\sqrt{\frac{r_S - r}{r_S}}\exp\left(\frac{r}{2r_S}\right)\cosh\left(\frac{t}{2r_S}\right). \tag{7.116}$$

We should interpret $-t$ as the Schwarzschild time in the second outer Schwarzschild region III ($u < -|v|$), since v is a forward evolving time parameter. The lower $r = 0$ singularity is then interpreted as a *white hole*, since every timelike geodesic in the region IV ($v < -|u|$) has to leave that region.

The three-dimensional spatial hyperplane $v = 0$ has $r \geqslant r_S$ everywhere and corresponds to a throat geometry that arises from gluing together two funnels from figure 7.1. This is shown in figure 7.6 along with the auxiliary embedding $Z^2 = 4r_S(r - r_S)$ of the glued together exterior regions of the Schwarzschild $\{x, y\}$ planes, see equation (7.50).

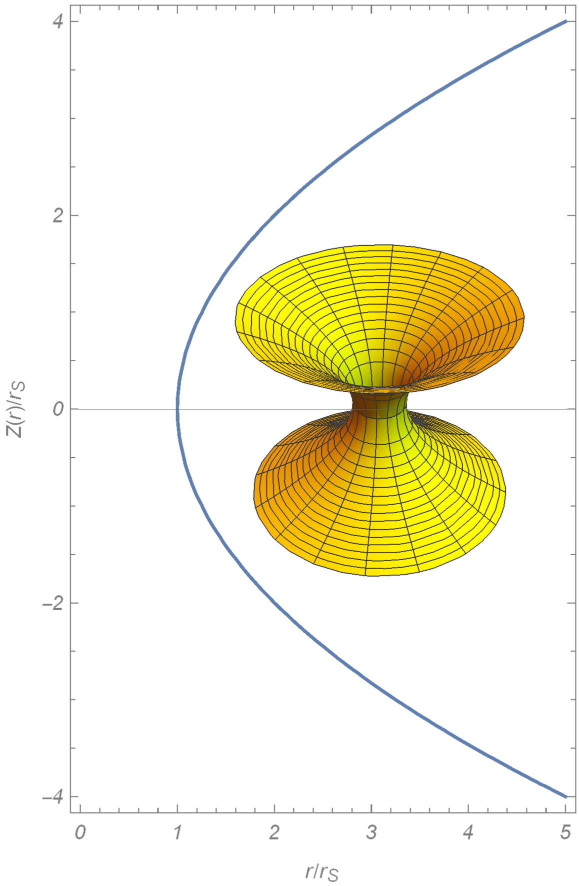

Figure 7.6. The embeddings of exterior regions of the $\{x, y\}$ planes of two Schwarzschild spacetimes in an auxiliary three-dimensional flat space.

For $v \neq 0$ the spatial geometry along v = const. still resembles the throat geometry, but now with three-dimensional throats (spacelike three-dimensional cylinder hypersurfaces with first decreasing and then again increasing spherical surface elements $r^2(d\vartheta^2 + \sin^2\vartheta\, d\varphi^2)$) connecting two Schwarzschild spheres. Therefore this spacetime is denoted as a wormhole spacetime.

Note that going through the throat at time $v = 0$ from one outer $\{x, y\}$ region into the second outer $\{x, y\}$ region is very different from the situation visualized in figure 7.1, which concerns going from an outer $\{x, y\}$ region into the black hole. Indeed, going through the throat along $v = 0$ touches the boundaries $r = r_S$ of both the black hole and the white hole. We can also think of timelike cones like the one in the lower half of figure 7.1 attached at the throat radius $r = r_S$. If we continue along $v = 0$, we cross e.g. from the upper part of the throat into the lower part of the throat, whereas a turn to $v > 0$ takes us into the black hole.

Of course going through the wormhole for constant time v would require infinite speed $|du/dv| = \infty$. This already indicates that going through the wormhole does not work, because we also have to move forward in time at subluminal speeds $|du/dv| < 1$. Any attempt to go through the wormhole necessarily involves penetrating the Schwarzschild horizon $t = \infty$, $r = r_S$ surrounding the black hole singularity. Inside the horizon we will again hit the singularity at time $r = 0$ and suffer the same consequences as going into an ordinary black hole.

Chapter 8

Massive particles in the Schwarzschild spacetime

To discuss the impact of the Schwarzschild geometry on the motion of free particles, we need to solve the geodesic equation (6.16) in the Schwarzschild spacetime, and we will use the Schwarzschild coordinates (7.47). We have noticed that the Schwarzschild coordinates are singular at the horizon $r = r_S$, and therefore we have to be careful when interpreting effects of horizon crossing in these coordinates. On the other hand, Schwarzschild coordinates are more intuitive for the description of motion outside of the horizon, because constant r implies constant distance $\ell(r, r_S)$ (7.49) from the horizon, and because t is (up to a scale factor) the time measured by an observer at constant distance from the horizon.

8.1 Massive particles in *t*-independent radially symmetric spacetimes

Before specializing to the actual Schwarzschild metric, we will start with a general t-independent rotationally symmetric metric

$$ds^2 = -c^2B(r)dt^2 + A(r)dr^2 + r^2d\vartheta^2 + r^2\sin^2\vartheta\, d\varphi^2. \quad (8.1)$$

The corresponding eigentime Lagrange function (6.23),

$$\frac{2}{m}L_\tau = -c^2B(r)\left(\frac{dt}{d\tau}\right)^2 + A(r)\left(\frac{dr}{d\tau}\right)^2 + r^2\left(\frac{d\vartheta}{d\tau}\right)^2 + r^2\sin^2\vartheta\left(\frac{d\varphi}{d\tau}\right)^2, \quad (8.2)$$

yields the equations of motion of a massive particle in terms of its eigentime $d\tau^2 = -ds^2/c^2$ from the Euler–Lagrange equations,

$$\frac{d^2r}{d\tau^2} + \frac{A'}{2A}\left(\frac{dr}{d\tau}\right)^2 - \frac{r}{A}\left(\frac{d\vartheta}{d\tau}\right)^2 - \frac{r\sin^2\vartheta}{A}\left(\frac{d\varphi}{d\tau}\right)^2 + \frac{c^2B'}{2A}\left(\frac{dt}{d\tau}\right)^2 = 0, \quad (8.3)$$

doi:10.1088/2053-2571/aaf173ch8

$$\frac{d^2\vartheta}{d\tau^2} + \frac{2}{r}\frac{d\vartheta}{d\tau}\frac{dr}{d\tau} - \sin\vartheta\cos\vartheta\left(\frac{d\varphi}{d\tau}\right)^2 = 0, \tag{8.4}$$

$$\frac{d^2\varphi}{d\tau^2} + \frac{2}{r}\frac{d\varphi}{d\tau}\frac{dr}{d\tau} + 2\cot\vartheta\,\frac{d\varphi}{d\tau}\frac{d\vartheta}{d\tau} = 0, \tag{8.5}$$

$$\frac{d^2t}{d\tau^2} + \frac{B'}{B}\frac{dr}{d\tau}\frac{dt}{d\tau} = 0. \tag{8.6}$$

Like in Newtonian mechanics we still discuss the spatial aspects of the motion in a gravitational field in terms of coordinates r, ϑ and φ or corresponding coordinates x, y and z, and the location of the particle is still described by a vector $\boldsymbol{r} = r(\sin\vartheta\cos\varphi\,\boldsymbol{e}_x + \sin\vartheta\sin\varphi\,\boldsymbol{e}_y + \cos\vartheta\,\boldsymbol{e}_z) = r\boldsymbol{e}_r$ with a constant Cartesian basis $\{\boldsymbol{e}_x, \boldsymbol{e}_y, \boldsymbol{e}_z\}$, the only difference being that r is *not* the physical distance between the moving particle and the origin of our coordinate grid any more, and the spatial part of the metric is *not* given by the Cartesian scalar products: $g_{ij} \neq \boldsymbol{e}_i \cdot \boldsymbol{e}_j = \delta_{ij}$.

Nevertheless, the decomposition of acceleration in radial and angular directions proceeds in the same way as in classical mechanics. The standard (in the Cartesian sense normalized) tangent vectors in radial and angular directions are

$$\boldsymbol{e}_r = \partial_r\boldsymbol{r} = \sin\vartheta\cos\varphi\,\boldsymbol{e}_x + \sin\vartheta\sin\varphi\,\boldsymbol{e}_y + \cos\vartheta\,\boldsymbol{e}_z, \tag{8.7}$$

$$\boldsymbol{e}_\vartheta = \frac{1}{r}\partial_\vartheta\boldsymbol{r} = \cos\vartheta\cos\varphi\,\boldsymbol{e}_x + \cos\vartheta\sin\varphi\,\boldsymbol{e}_y - \sin\vartheta\,\boldsymbol{e}_z, \tag{8.8}$$

$$\boldsymbol{e}_\varphi = \frac{1}{r\sin\vartheta}\partial_\varphi\boldsymbol{r} = -\sin\varphi\,\boldsymbol{e}_x + \cos\varphi\,\boldsymbol{e}_y, \tag{8.9}$$

and these equations imply

$$\frac{d}{d\tau}\boldsymbol{e}_r = \dot\vartheta\,\boldsymbol{e}_\vartheta + \dot\varphi\sin\vartheta\,\boldsymbol{e}_\varphi, \quad \frac{d}{d\tau}\boldsymbol{e}_\vartheta = -\dot\vartheta\,\boldsymbol{e}_r + \dot\varphi\cos\vartheta\,\boldsymbol{e}_\varphi, \tag{8.10}$$

$$\frac{d}{d\tau}\boldsymbol{e}_\varphi = -\dot\varphi(\sin\vartheta\,\boldsymbol{e}_r + \cos\vartheta\,\boldsymbol{e}_\vartheta). \tag{8.11}$$

This yields the following equations for velocity and acceleration in polar coordinates[1]:

$$\dot{\boldsymbol{r}} = \dot r\boldsymbol{e}_r + r\dot\vartheta\boldsymbol{e}_\vartheta + r\dot\varphi\sin\vartheta\,\boldsymbol{e}_\varphi, \tag{8.12}$$

$$\begin{aligned}\ddot{\boldsymbol{r}} = {} & (\ddot r - r\dot\vartheta^2 - r\dot\varphi^2\sin^2\vartheta)\boldsymbol{e}_r + (2\dot r\dot\vartheta + r\ddot\vartheta - r\dot\varphi^2\sin\vartheta\cos\vartheta)\boldsymbol{e}_\vartheta \\ & + (2\dot r\dot\varphi\sin\vartheta + r\ddot\varphi\sin\vartheta + 2r\dot\varphi\dot\vartheta\cos\vartheta)\boldsymbol{e}_\varphi.\end{aligned} \tag{8.13}$$

[1] These equations are the same as in mechanics, the only difference being that here derivatives are taken with respect to the eigentime τ instead of the coordinate time t.

If we now compare the expressions for the accelerations in the angular directions with the equations (8.4) and (8.5), we recognize that the angular accelerations vanish. Furthermore, substitution of $\ddot{r}$ from equation (8.3) into (8.13) yields

$$\ddot{\boldsymbol{r}} = -\left[\frac{A'}{2A}\dot{r}^2 - \frac{r}{A}\dot{\vartheta}^2 - \frac{r\sin^2\vartheta}{A}\dot{\varphi}^2 + \frac{c^2B'}{2A}\dot{t}^2 + r\dot{\vartheta}^2 + r\dot{\varphi}^2\sin^2\vartheta\right]\boldsymbol{e}_r, \tag{8.14}$$

i.e. gravity around a non-rotating spherically symmetric mass distribution is still described by a central force: $\ddot{\boldsymbol{r}}\|\boldsymbol{r}$, and therefore the angular momentum or 'area velocity' (but this does not measure the rate of change of a physical area any more!)

$$\frac{1}{2}\boldsymbol{J} = \frac{1}{2}\boldsymbol{r}\times\dot{\boldsymbol{r}} = \frac{\boldsymbol{L}}{2m} \tag{8.15}$$

is still conserved and the motion takes place in a plane $\perp\boldsymbol{J}$, which we can choose as the plane $\vartheta = \pi/2$. The equations of motion (8.3)–(8.6) thus reduce to

$$\frac{d^2r}{d\tau^2} + \frac{A'}{2A}\left(\frac{dr}{d\tau}\right)^2 - \frac{r}{A}\left(\frac{d\varphi}{d\tau}\right)^2 + \frac{c^2B'}{2A}\left(\frac{dt}{d\tau}\right)^2 = 0, \tag{8.16}$$

$$\frac{d^2\varphi}{d\tau^2} + \frac{2}{r}\frac{d\varphi}{d\tau}\frac{dr}{d\tau} = \frac{1}{r^2}\frac{d}{d\tau}(r^2\dot{\varphi}) = 0, \tag{8.17}$$

$$\frac{d^2t}{d\tau^2} + \frac{B'}{B}\frac{dr}{d\tau}\frac{dt}{d\tau} = \frac{1}{B}\frac{d}{d\tau}(B\dot{t}) = 0, \tag{8.18}$$

and we have found two integrals of the motion (actually four integrals if we include the constant orientation of $\boldsymbol{J}$):

$$\dot{\varphi}(\tau) = \frac{J}{r^2(\tau)}, \qquad \dot{t}(\tau) = \frac{\gamma}{B(r(\tau))}. \tag{8.19}$$

The equation (8.15) implies that the constant J is the related to the conserved magnitude of angular momentum, $J \equiv |\boldsymbol{J}| = |\boldsymbol{L}|/m$. The constant γ is related to energy conservation. The metric (8.1) is invariant under time translations $\delta x^0 = \text{const.}$, and therefore the results of section 6.4 and application of equation (6.100) yield a conserved energy

$$E = -mcg_{0\mu}\dot{x}^\mu = mc^2B(r(\tau))\dot{t}(\tau). \tag{8.20}$$

Comparison with the second equation in (8.19) shows that

$$\gamma = E/mc^2 \tag{8.21}$$

is the ratio between total energy and rest mass energy of the particle. I denoted this by γ since in STR we also have a relation $E = \gamma mc^2$. Note also that E is conserved everywhere in the Schwarzschild spacetime. However, t is a spacelike variable inside the horizon, and therefore $E/c = \gamma mc$ is rather a conserved spatial momentum for $r < r_S$.

Substitution of the equations (8.19) into equation (8.16) yields

$$\frac{d^2r}{d\tau^2} + \frac{A'}{2A}\left(\frac{dr}{d\tau}\right)^2 - \frac{J^2}{Ar^3} + \frac{c^2\gamma^2 B'}{2AB^2} = 0, \tag{8.22}$$

from which we find another first integral through multiplication with $2A\dot{r}$,

$$A\left(\frac{dr}{d\tau}\right)^2 + \frac{J^2}{r^2} - \frac{c^2\gamma^2}{B} = K. \tag{8.23}$$

However, contrary to J and γ, K turns out to be the same universal constant for all trajectories in the metric (8.1). To see this recall equation (8.1) for $\vartheta = \pi/2$:

$$ds^2 = -c^2 d\tau^2 = -c^2 B(r)dt^2 + A(r)dr^2 + r^2 d\varphi^2. \tag{8.24}$$

This yields with the equations (8.19) the equation

$$A\left(\frac{dr}{d\tau}\right)^2 + \frac{J^2}{r^2} - \frac{c^2\gamma^2}{B} = A\left(\frac{dr}{d\tau}\right)^2 + r^2\left(\frac{d\varphi}{d\tau}\right)^2 - c^2 B\left(\frac{dt}{d\tau}\right)^2 = -c^2, \tag{8.25}$$

and comparison with equation (8.23) yields $K = -c^2$.

We have thus found that particles in a spherically symmetric gravitational field move in a plane subject to the following three equations:

$$\dot{\varphi}(\tau) = \frac{J}{r^2(\tau)}, \tag{8.26}$$

$$\dot{t}(\tau) = \frac{\gamma}{B(r(\tau))}, \tag{8.27}$$

and

$$\frac{1}{2}\left(\frac{dr}{d\tau}\right)^2 + \frac{J^2}{2Ar^2} + \frac{c^2}{2A} - \frac{c^2\gamma^2}{2AB} = 0, \tag{8.28}$$

where the result $K = -c^2$ was inserted. Equation (8.28) yields in the Schwarzschild metric the radial equation of motion

$$\frac{1}{2}\left(\frac{dr}{d\tau}\right)^2 = \frac{c^2\gamma^2}{2} - \frac{r - r_S}{2r}\left(c^2 + \frac{J^2}{r^2}\right). \tag{8.29}$$

8.2 Radial motion in terms of the effective potential

We now specialize our previous discussion of motion in general t-independent and spherically symmetric spacetimes to the Schwarzschild solution $A^{-1}(r) = B(r) = (r - r_S)/r$.

In a mechanical analog of equation (8.29), the radial motion of a massive particle in terms of its eigentime corresponds to the motion of a non-relativistic particle of specific energy (i.e. energy per mass) $\gamma^2 c^2/2$ and specific effective potential (i.e. again normalized to particle mass)

$$\tilde{V}(r) = \frac{r - r_S}{2r}\left(c^2 + \frac{J^2}{r^2}\right) = \left(\frac{1}{2} - \frac{GM}{c^2 r}\right)\left(c^2 + \frac{J^2}{r^2}\right). \tag{8.30}$$

Note that

$$0 \leqslant \frac{1}{2}\left(\frac{dr}{d\tau}\right)^2 = \frac{c^2\gamma^2}{2} - \frac{r - r_S}{2r}\left(c^2 + \frac{J^2}{r^2}\right) = \frac{c^2\gamma^2}{2} - \tilde{V}(r), \tag{8.31}$$

and therefore only those values of r are accessible to the particle which satisfy

$$\tilde{V}(r) = \frac{r - r_S}{2r}\left(c^2 + \frac{J^2}{r^2}\right) \leqslant \frac{c^2\gamma^2}{2}. \tag{8.32}$$

In particular, $\tilde{V}(r \to \infty) = c^2/2$, i.e. a particle coming from infinity or going to infinity must have $\gamma \geqslant 1$.

The effective potential (8.30) can be written with $j \equiv J/cr_S$ and $x \equiv r/r_S$ in the form

$$\frac{2}{c^2}\tilde{V}(r) = 1 - \frac{r_S}{r} + \frac{J^2}{c^2 r^2} - \frac{J^2 r_S}{c^2 r^3} = 1 - \frac{1}{x} + \frac{j^2}{x^2} - \frac{j^2}{x^3}, \tag{8.33}$$

and the derivative is

$$\begin{aligned}\frac{2}{c^2}r^4\frac{d\tilde{V}}{dr} &= r_S r^2 - \frac{2J^2}{c^2}r + \frac{3J^2 r_S}{c^2} = r_S^3(x^2 - 2j^2 x + 3j^2)\\ &= r_S^3[(x - j^2)^2 + j^2(3 - j^2)].\end{aligned} \tag{8.34}$$

The derivative $\tilde{V}'(r)$ is positive semidefinite for

$$j^2 = \frac{J^2}{c^2 r_S^2} \leqslant 3, \tag{8.35}$$

and the effective potential for the radial motion is attractive everywhere, i.e. there is no centrifugal barrier in this case and no bound orbits can exist.

On the other hand, for

$$j^2 > 3 \quad (\Rightarrow \quad J > \sqrt{3}\,cr_S) \tag{8.36}$$

the potential has a maximum at

$$r_{\max} = r_S\left(j^2 - j\sqrt{j^2 - 3}\right). \tag{8.37}$$

The location of the maximum decreases with increasing $j^2 > 3$ with limiting values

$$r_{\max}|_{j^2 \to 3} = 3r_S > r_{\max} \geqslant r_{\max}|_{j^2 \to \infty} = \frac{3}{2}r_S, \tag{8.38}$$

Furthermore, the potential for $j^2 > 3$ also has a minimum at

$$r_{\min} = r_S\left(j^2 + j\sqrt{j^2 - 3}\right) > 3r_S > r_{\max}. \tag{8.39}$$

The maximal height $\tilde{V}(r_{\max})$ of the centrifugal barrier increases with increasing j,

$$\begin{aligned} 3 < j^2 < 4 \quad &\Leftrightarrow \quad 4c^2/9 < \tilde{V}(r_{\max}) < c^2/2, \\ j^2 \geqslant 4 \quad &\Leftrightarrow \quad \tilde{V}(r_{\max}) \geqslant c^2/2. \end{aligned} \tag{8.40}$$

The local minima for sufficiently large angular momentum appear extremely shallow in figure 8.1, since we use an extremely large unit c^2 for the potential and an extremely small unit r_S for the radial coordinate.

For the **discussion of the various cases** we note that the different kinds of orbits depend on the angular momentum $L = mJ = mjcr_S$ and the energy $E = \gamma mc^2$ of the particle, and sometimes also on the initial values $r(0)$ and $\dot{r}(0)$ of the radial coordinate and radial speed. We will enumerate the different possible cases first by increasing angular momentum j and then by increasing energy γ.

1. The angular momentum is too small to generate a centrifugal barrier, i.e. $j^2 \leqslant 3$.
   a. $j^2 \leqslant 3$ and $\gamma^2 < 1$:
      The particle can reach at most a finite maximal radius r_2 given by

$$\tilde{V}(r_2) = \frac{c^2\gamma^2}{2}. \tag{8.41}$$

      Finally it will always collide with the central mass or disappear into the black hole, respectively.
   b. $j^2 \leqslant 3$ and $\gamma^2 \geqslant 1$:
      The particle will escape to $r \to \infty$ if $\dot{r}(0) > 0$. Otherwise it will collide with the central mass or disappear into the black hole, respectively.

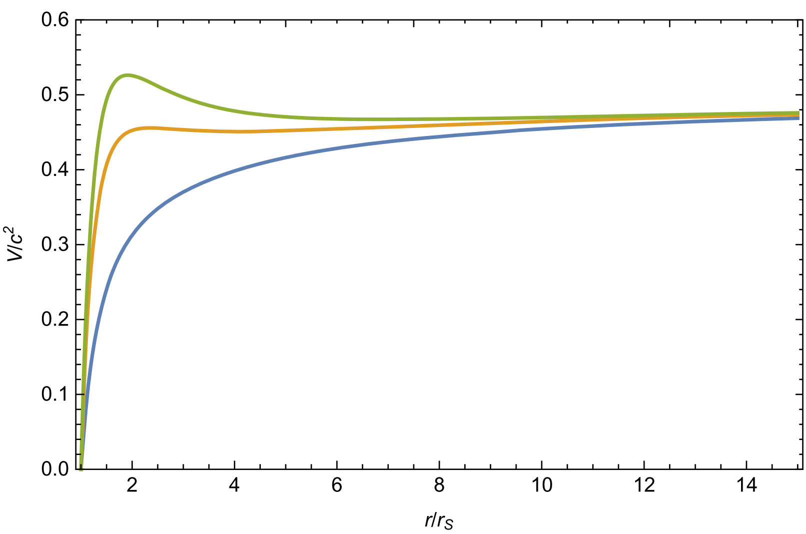

Figure 8.1. The effective potential $\tilde{V}/c^2$ for $r_S \leqslant r \leqslant 15r_S$ and $j = 1$ (blue), $j = 1.8$ (brown), $j = 2.1$ (green).

2. The angular momentum generates a centrifugal barrier with height $\tilde{V}(r_{\max}) < c^2/2$, i.e. $3 < j^2 < 4$.
   a. $3 < j^2 < 4$, $\gamma^2 < 2\tilde{V}(r_{\max})/c^2$ and $r(0)$ is close to the Schwarzschild radius (i.e. $r(0)$ is on the left of the centrifugal barrier):
   The particle can reach at most a finite maximal radius r_2 given by

$$\tilde{V}(r_2) = \frac{c^2\gamma^2}{2} \tag{8.42}$$

   and will finally collide with the central mass or disappear into the black hole, respectively.
   b. $3 < j^2 < 4$,

$$\frac{2}{c^2}\tilde{V}(r_{\min}) \leqslant \gamma^2 < \frac{2}{c^2}\tilde{V}(r_{\max}), \tag{8.43}$$

   and the initial radius $r(0)$ is to the right of the centrifugal barrier:
   There exist collision free bound orbits with $r_1 \leqslant r \leqslant r_2$, where

$$\tilde{V}(r_1) = \tilde{V}(r_2) = \frac{c^2\gamma^2}{2}. \tag{8.44}$$

   c. $3 < j^2 < 4$ and

$$\frac{2}{c^2}\tilde{V}(r_{\max}) \leqslant \gamma^2 < 1. \tag{8.45}$$

   Like in case **2a**, the particle can reach at most a finite maximal radius r_2 given by

$$\tilde{V}(r_2) = \frac{c^2\gamma^2}{2} \tag{8.46}$$

   and will finally collide with the central mass or disappear into the black hole, respectively. However, this time this happens because the centrifugal barrier is too small for the particle's energy to stop the infall.
   d. $3 < j^2 < 4$ and $\gamma^2 \geqslant 1$:
   The particle will escape to $r \to \infty$ if $\dot{r}(0) > 0$. Otherwise it will collide with the central mass or disappear into the black hole, respectively.

3. The angular momentum generates a centrifugal barrier with height $\tilde{V}(r_{\max}) \geqslant c^2/2$, i.e. $j^2 \geqslant 4$.
   a. $j^2 \geqslant 4$,

$$\gamma^2 < \frac{2}{c^2}\tilde{V}(r_{\max}), \tag{8.47}$$

   and $r(0)$ is on the left of the centrifugal barrier:

The particle again can reach at most a finite maximal radius r_2 given by

$$\tilde{V}(r_2) = \frac{c^2\gamma^2}{2} \tag{8.48}$$

and will finally collide with the central mass or disappear into the black hole, respectively.

b. $j^2 \geqslant 4$,

$$\frac{2}{c^2}\tilde{V}(r_{\min}) \leqslant \gamma^2 < 1 \tag{8.49}$$

and $r(0)$ is on the right of the centrifugal barrier:

There exist collision free bound orbits with $r_1 \leqslant r \leqslant r_2$, where

$$\tilde{V}(r_1) = \tilde{V}(r_2) = \frac{c^2\gamma^2}{2}. \tag{8.50}$$

c. $j^2 \geqslant 4$,

$$1 \leqslant \gamma^2 < \frac{2}{c^2}\tilde{V}(r_{\max}), \tag{8.51}$$

and $r(0)$ is on the right of the centrifugal barrier:

The particle can reach at most a minimal radius r_1 given by

$$\tilde{V}(r_1) = \frac{c^2\gamma^2}{2} \tag{8.52}$$

and will finally escape to $r \to \infty$.

d. $j^2 \geqslant 4$ and

$$\frac{2}{c^2}\tilde{V}(r_{\max}) < \gamma^2. \tag{8.53}$$

The particle will escape to $r \to \infty$ if $\dot{r}(0) > 0$. Otherwise it will collide with the central mass or disappear into the black hole, respectively.

The Earth's motion around the Sun corresponds to the case **3b**. For **estimates of the parameters j and γ for Earth's orbit** we notice that the radius $r_\oplus = 1.496 \times 10^8$ km of the Earth's orbit is much larger than the Schwarzschild radius $r_{\odot S} = 2GM_\odot/c^2 = 2.953$ km of the Sun. Therefore the last term in the effective potential for the Earth's motion is negligible,

$$\begin{aligned}\tilde{V}(r_\oplus) &= \left(\frac{1}{2} - \frac{GM_\odot}{c^2 r_\oplus}\right)\left(c^2 + \frac{J^2}{r_\oplus^2}\right) = -\frac{GM_\odot}{r_\oplus} + \frac{J^2}{2r_\oplus^2} + \frac{c^2}{2} - \frac{J^2 r_{\odot S}}{2r_\oplus^3} \\ &\simeq -\frac{GM_\odot}{r_\oplus} + \frac{J^2}{2r_\oplus^2} + \frac{c^2}{2}.\end{aligned} \tag{8.54}$$

Furthermore, the Earth's orbit is approximately circular, which implies

$$\tilde{V}'(r_\oplus) \approx 0 \quad \Rightarrow \quad J^2 \approx GM_\odot r_\oplus = \frac{c^2}{2} r_{\odot S} r_\oplus \tag{8.55}$$

and therefore

$$j = \frac{J}{c r_{\odot S}} \approx \sqrt{\frac{r_\oplus}{2 r_{\odot S}}} = 5 \times 10^3. \tag{8.56}$$

The minimum in the effective potential of the Earth is approximately

$$\begin{aligned} \tilde{V}(r_\oplus) - \tilde{V}(\infty) = \tilde{V}(r_\oplus) - \frac{c^2}{2} &\approx -\frac{GM_\odot}{r_\oplus} + \frac{J^2}{2r_\oplus^2} = -\frac{GM_\odot}{2r_\oplus} = -\frac{r_{\odot S}}{4r_\oplus}c^2 \\ &= -5 \times 10^{-9} c^2. \end{aligned} \tag{8.57}$$

The γ parameter for the approximately circular orbit of the Earth is

$$\gamma_\oplus^2 \approx \frac{2}{c^2} \tilde{V}(r_\oplus) \approx 1 - 10^{-8}. \tag{8.58}$$

Limits on closest approaches in the previous cases 2b, 3b and 3c

We assume that r_S is located *outside* of the mass distribution which generates the Schwarzschild metric, i.e. we are talking about a *Schwarzschild black hole*. Freely falling particles have to be reflected at the centrifugal barrier in order not to collide with the black hole. This limits the closest approaches which can be reached with force free motion (if we can use propulsion, the limit of closest approach without falling into the black hole is $r_S + \epsilon$):

The location r_{max} of the tip of the centrifugal barrier decreases with increasing j, see equation (8.38),

$$\lim_{j\to\infty} r_{max} = \frac{3}{2} r_S, \tag{8.59}$$

i.e. there cannot be any collision free orbit with a periastron closer than[2] $1.5r_S = 3GM/c^2$. The non-collision orbits which approximate the minimal periastron value $1.5r_S$ correspond to satellites falling towards the black hole from $r \to \infty$ and eventually escaping again to $r \to \infty$ after several revolutions around the black hole (because for those particles $r'(\varphi) = \dot{r}(\tau)/\dot{\varphi}(\tau)$ is small near their periastron). Note that the particles approaching that point of closest approach in free fall and then eventually escaping again must have both $j \to \infty$ and $\gamma \to \infty$, because the condition in the point of closest approach is $\tilde{V}(r_{max}) = c^2\gamma^2/2$. The asymptotic value of $\tilde{V}(r_{max})$ for $j \gg 1$ is

[2] This result was derived for the first time by Darwin as late as 1959—43 years after discovery of the Schwarzschild metric! See Darwin C 1959 *Proc. R. Soc. London* A249 180.

$$\tilde{V}(r_{\max})|_{j\gg 1} \simeq \frac{c^2}{6}\left(1+\frac{4}{9}j^2\right) \simeq \frac{2c^2}{27}j^2, \tag{8.60}$$

and therefore γ must be tuned with j according to

$$\gamma \simeq 2j/\sqrt{27} \gg 1 \tag{8.61}$$

to approximate the closest periastron value (8.59) in free fall.

To actually *reach* the minimal periastron value (8.59) requires $\sqrt{27}\gamma = 2j \to \infty$. This implies with equations (8.26) and (8.29) the equation $r'(\varphi)|_{r=3r_S/2} = 0$, i.e. in this case the particle would actually have to orbit the black hole on the circle $r = 3r_S/2$. However, since $\gamma = E/mc^2 \to \infty$ is required, this limiting case could only be attained by massless particles. We also note that this orbit is unstable. Any fluctuation to larger or smaller radius would lead to escape to $r \to \infty$ or collision with the central black hole. We will rediscover this extreme solution in our discussion of motion of massless particles in the Schwarzschild spacetime.

The radius of closest approach corresponds to a distance from the Schwarzschild horizon

$$\begin{aligned}\ell &= \int_{r_S}^{1.5r_S} dr\sqrt{\frac{r}{r-r_S}} = \left[\sqrt{r(r-r_S)} + r_S\ln\left(\sqrt{r}+\sqrt{r-r_S}\right)\right]_{r_S}^{1.5r_S} \\ &= r_S\frac{\sqrt{3}}{2} + r_S\ln\left(\frac{\sqrt{3}+1}{\sqrt{2}}\right) \simeq 1.5245r_S.\end{aligned} \tag{8.62}$$

We can get an even sharper lower bound on the periastron of collision free *stable bound* free fall orbits. For stable bound orbits the energy parameter γ is restricted by $\gamma < 1$ to ensure that the particle is reflected at the effective potential for some maximal radius r_2. The limit on the energy implies that our particle will be deflected on the right shoulder, $r_1 > r_{\max}$, of the centrifugal barrier, if j is so large that $\tilde{V}(r_{\max}) > c^2/2$. Therefore, for bound orbits the closest periastron is not given by the global minimal possible value $1.5r_S$ of $r_{\max}$, but by the minimal possible value of $r_{\max}$ which can be reached in a bound orbit. Since $r_{\max}$ decreases with j while $\tilde{V}(r_{\max})$ increases with j, we want to have maximal j ($\Rightarrow$ maximal $\tilde{V}(r_{\max})$) such that the bound particle can still reach the top $r_{\max}$ of the centrifugal barrier and is not deflected before that. Therefore the closest periastron that we can find for a collision free bound orbit corresponds to the case when the tip of the centrifugal barrier $\tilde{V}(r_{\max})$ just equals $\tilde{V}(\infty) = c^2/2$. The requirement $\tilde{V}(r_{\max}) = c^2/2$ yields with equations (8.37) and (8.33) the condition

$$\frac{r_{\max}}{r_S} = j^2 - j\sqrt{j^2-3} = \frac{j^2}{2} \pm \frac{j}{2}\sqrt{j^2-4}. \tag{8.63}$$

This implies

$$j = 2\sqrt{j^2-3} \pm \sqrt{j^2-4} \tag{8.64}$$

with solution $j = 2$. This yields for the periastron of *stable bound* free fall orbits the lower bound

$$r_{\text{periastron, bound}} \geqslant r_{\max}|_{j=2} = 2r_S = 4GM/c^2. \tag{8.65}$$

This is the minimal possible periastron value for eccentric stable orbits[3] around the black hole. *Stable circular* free fall orbits satisfy $r_{\text{circ}} = r_{\min}$. Equation (8.39) therefore shows that the radii of stable circular orbits are constrained by

$$r_{\text{circ}} > 3r_S. \tag{8.66}$$

8.3 The shape of the trajectory

The equations (8.26) and (8.29) yield with

$$\frac{dr}{d\varphi} = \frac{dr/d\tau}{d\varphi/d\tau} \tag{8.67}$$

the differential equation for the shape $r = r(\varphi)$ of the trajectory,

$$\left(\frac{dr}{d\varphi}\right)^2 + r(r - r_S)\left(1 + \frac{c^2}{J^2}r^2\right) = \frac{\gamma^2 c^2}{J^2}r^4, \tag{8.68}$$

or

$$\left(\frac{dr}{d\varphi}\right)^2 - \frac{c^2}{J^2}(\gamma^2 - 1)r^4 - \frac{c^2}{J^2}r_S r^3 + r^2 - r_S r = 0. \tag{8.69}$$

Substituting γ in terms of the conserved energy (8.20) of the particle,

$$\gamma = \frac{E}{mc^2} = 1 + \frac{E_{\text{class}}}{mc^2} \Rightarrow \gamma^2 - 1 = \frac{2E_{\text{class}}}{mc^2}\left(1 + \frac{E_{\text{class}}}{2mc^2}\right), \tag{8.70}$$

yields

$$\left(\frac{dr}{d\varphi}\right)^2 - \frac{2E_{\text{class}}}{mJ^2}\left(1 + \frac{E_{\text{class}}}{2mc^2}\right)r^4 - \frac{c^2}{J^2}r_S r^3 + r^2 - r_S r = 0, \tag{8.71}$$

and comparing this with the corresponding equation for the Kepler trajectories in Newtonian mechanics[4],

$$\left(\frac{dr}{d\varphi}\right)^2 - \frac{2E_{\text{class}}}{mJ^2}r^4 - \frac{c^2}{J^2}r_S r^3 + r^2 = 0, \tag{8.72}$$

shows that the effects of general relativity is a modification of the coefficient in the r^4 term and the introduction of an additional term $-r_S r$ in the equation for the trajectory. The Newtonian limit thus corresponds to $r \gg r_S$ and $|E_{\text{class}}| \ll mc^2$, i.e. Newtonian gravity is a good approximation as long as we are far away from the

[3] These are like elliptic orbits in Newtonian mechanics except for the periastron precession; see the discussion below on shapes of orbits.

[4] See e.g. Goldstein H, Poole C and Safko J 2002 *Classical Mechanics* 3rd edn (San Francisco, CA: Addison-Wesley), or any other mechanics textbook.

Schwarzschild horizon, and as long as the kinetic plus gravitational energy of the moving particle (planet, star, satellite) is small compared to its rest energy.

Separation of variables in equation (8.69) yields φ as an elliptic integral:

$$\pm(\varphi - \varphi_0) = \int dr \frac{J}{\sqrt{r(c^2(\gamma^2 - 1)r^3 + c^2 r_S r^2 - J^2 r + J^2 r_S)}}. \tag{8.73}$$

If not for the last term under the square root, this would be the same solvable integral as in Newtonian mechanics, only with $2E_{\text{class}}/m$ replaced with $c^2(\gamma^2 - 1)$. Therefore, neglecting the last term we would again find ellipses for $\gamma < 1 \Rightarrow E < mc^2$, parabolas for $\gamma = 1 \Rightarrow E = mc^2$, and hyperbolas for $\gamma > 1 \Rightarrow E > mc^2$. However, the last term changes this picture even for $r \gg r_S$ in a minute, but observable way:

Periastron precession

We have seen that there exist bound orbits for

$$\frac{2}{c^2}\tilde{V}(r_{\min}) \leqslant \gamma^2 < 1. \tag{8.74}$$

To discuss the impact of the relativistic correction term r_{Sr} on these orbits we expand the elliptic integral in equation (8.73) in this term[5]:

$$\begin{aligned}
\pm(\varphi - \varphi_0) &= \int dr \frac{J}{\sqrt{r(c^2(\gamma^2 - 1)r^3 + c^2 r_S r^2 - J^2 r + J^2 r_S)}} \\
&\approx \int dr \frac{J}{r\sqrt{c^2(\gamma^2 - 1)r^2 + c^2 r_S r - J^2}} \\
&\quad - \int dr \frac{J^3 r_S}{2r^2\sqrt{c^2(\gamma^2 - 1)r^2 + c^2 r_S r - J^2}^3} \\
&= -\arccos\left(\frac{c^2 r_S r - 2J^2}{r\sqrt{c^4 r_S^2 + 4c^2(\gamma^2 - 1)J^2}}\right)\left(1 + \frac{3c^2 r_S^2}{4J^2}\right) \\
&\quad + \frac{J}{2\sqrt{c^2(\gamma^2 - 1)r^2 + c^2 r_S r - J^2}} \\
&\quad \times \left(\frac{c^2(\gamma^2 - 1)r_S}{J^2}\frac{8(\gamma^2 - 1)J^2 + 3c^2 r_S^2}{4(\gamma^2 - 1)J^2 + c^2 r_S^2} r\right. \\
&\quad \left. + \frac{c^2 r_S^2}{J^2}\frac{10(\gamma^2 - 1)J^2 + 3c^2 r_S^2}{4(\gamma^2 - 1)J^2 + c^2 r_S^2} - \frac{r_S}{r}\right),
\end{aligned} \tag{8.75}$$

[5] We can do this only for ellipses of sufficient eccentricity and far away from the periastron and apastron: In the periastron and apastron $r'(\varphi) = 0$, which implies that there the sum of zeroth-order terms in the square root in equation (8.73) equals the first-order relativistic correction, whence the expansion of the square root is not possible there. Therefore equation (8.75) holds only for the change of φ along pieces of the orbit away from the periastron and apastron, and we actually calculate the condition for the increase of the angle until we can reach again a point between the periastron and apastron. However, this also implies a corresponding angular shift for the location of the periastron.

or keeping again terms of order r_S/r

$$\frac{c^2 r_S r - 2J^2}{r\sqrt{c^4 r_S^2 + 4c^2(\gamma^2 - 1)J^2}} = \cos\Bigg[\frac{J}{2\sqrt{c^2(\gamma^2 - 1)r^2 + c^2 r_S r - J^2}}$$
$$\times\left(\frac{c^2(\gamma^2 - 1)r_S}{J^2}\frac{8(\gamma^2 - 1)J^2 + 3c^2 r_S^2}{4(\gamma^2 - 1)J^2 + c^2 r_S^2}r + \frac{c^2 r_S^2}{J^2}\frac{10(\gamma^2 - 1)J^2 + 3c^2 r_S^2}{4(\gamma^2 - 1)J^2 + c^2 r_S^2} - \frac{r_S}{r}\right) \quad (8.76)$$
$$\mp\left(1 - \frac{3c^2 r_S^2}{4J^2}\right)(\varphi - \varphi_0)\Bigg].$$

The factor $\left(1 + \frac{3c^2 r_S^2}{4J^2}\right)^{-1}$ multiplying φ implies that r can attain the same value again only after φ has increased by

$$\varphi \to \varphi + 2\pi + \frac{3c^2 r_S^2}{2J^2}\pi. \quad (8.77)$$

The additional increment

$$\Delta\varphi = \frac{3c^2 r_S^2}{2J^2}\pi = 6\pi\frac{G^2 M^2}{c^2 J^2} \quad (8.78)$$

in the angle is the *periastron advance* due to general relativity. For $j = J/(cr_S) \gg 1$ we find from

$$r_{\min} = r_S\left(j^2 + j\sqrt{j^2 - 3}\right) \approx 2j^2 r_S \quad (8.79)$$

$$J^2 \approx \frac{1}{2}c^2 r_S r_{\min} = GMr_{\min} \quad (8.80)$$

and

$$\Delta\varphi \approx 3\pi\frac{r_S}{r_{\min}} = 6\pi\frac{GM}{c^2 r_{\min}}. \quad (8.81)$$

The relativistic perihelion advance of the planets is largest for the innermost planets, and it is usually a small effect compared to the perihelion shift induced by the gravitational effects of the other planets. However, the observation of the relativistic effect in the perihelion advance of Mercury was one of the early observational confirmations of general relativity. The observed value for the precession of the perihelion of Mercury is $\Delta\varphi_{\text{observed}}/T_{\text{Mercury}} = 5600''$/century, while the effects of other planets account for $\Delta\varphi_{\text{Newton}}/T_{\text{Mercury}} = 5557''$/century. The discrepancy of $\Delta\varphi/T_{\text{Mercury}} = 43''$/century is in excellent agreement with the prediction (8.81) from general relativity.

8.4 Clocks in the Schwarzschild spacetime

Before determining the trajectories of the massive particles further, we will study the behavior of clocks in the Schwarzschild spacetime.

Equation (6.5) tells us that the time $d\tau$ which elapses on an ideal clock while it moves from $\{t, r, \vartheta, \varphi\}$ to $\{t + dt, r + dr, \vartheta + d\vartheta, \varphi + d\varphi\}$ is

$$d\tau = \frac{1}{c}\sqrt{\frac{r - r_S}{r}c^2dt^2 - \frac{r}{r - r_S}dr^2 - r^2d\vartheta^2 - r^2\sin^2\vartheta\, d\varphi^2}\,. \tag{8.82}$$

This tells us in particular that t is the time measured by ideal clocks which are at rest at $r \to \infty$. Equation (8.82) also tells us that every other ideal clock in the region $r > r_S$ measures a time

$$d\tau < dt, \tag{8.83}$$

i.e. every clock which is closer to the mass M ticks at a slower rate, and furthermore, every clock which is moving at a radial coordinate $r > r_S$ is further slowed down than a local clock which is at rest at the radial coordinate r,

$$d\tau|_r \leqslant dt_r \equiv \sqrt{\frac{r - r_S}{r}}\,dt < dt. \tag{8.84}$$

More specifically, equation (8.27) tells us that the time interval $d\tau$ of a freely falling clock at radial coordinate $r(\tau)$ is related to the time interval dt measured on a stationary clock at $r \to \infty$ through

$$dt = \frac{\gamma}{B(r(\tau))}d\tau = \gamma\frac{r(\tau)}{r(\tau) - r_S}d\tau. \tag{8.85}$$

If we compare this with the equation for time dilation of a moving clock showing time τ in STR,

$$dt = \frac{1}{\sqrt{1 - \beta^2}}d\tau = \gamma_\beta d\tau, \tag{8.86}$$

we recognize that $\gamma/B(r(\tau))$ is the time dilation factor for the freely falling particle at radial coordinate $r(\tau)$ compared to a stationary clock at $r \to \infty$. In particular, if $\gamma \geqslant 1$ then

$$1 \leqslant \gamma = \frac{c}{\sqrt{c^2 - v_\infty^2}} \tag{8.87}$$

is the time dilation factor at $r \to \infty$, where the particle has speed v_∞, with $\gamma = 1$ corresponding to $v_\infty = 0$. Furthermore, equation (8.32) implies with $r > r_S$,

$$\gamma\frac{r}{r - r_S} > \gamma\sqrt{\frac{r}{r - r_S}} \geqslant \sqrt{1 + \frac{J^2}{c^2r^2}} \geqslant 1, \tag{8.88}$$

and therefore $\gamma r(\tau)/(r(\tau) - r_S)$ and $\gamma\sqrt{r(\tau)/(r(\tau) - r_S)}$ are always dilation factors, even for $\gamma < 1$. Equation (8.84) relating dt_r and dt and equation (8.89) for the local freely falling clock imply

$$dt = \sqrt{\frac{r}{r - r_S}}dt_r > dt_r = \gamma\sqrt{\frac{r}{r - r_S}}d\tau \geqslant d\tau, \tag{8.89}$$

In a frame which is at rest relative to the black hole horizon, the freely falling clock ticks slower than a local clock at rest, which in turn ticks slower than the Schwarzschild clock t at infinity. We need to specify the rest frame in this formulation, because the effect of relative motion between $d\tau$ and dt_r is the special relativistic motion effect which is reciprocal between the local rest frame and the falling frame (remember that the freely falling observer has a different notion of simultaneity).

We note in particular that the clock which is at rest deeper in the gravitational field is slowed down relative to the clock farther out. This is the gravitational time dilation effect that we have discovered already in equation (6.47) for weak gravitational fields.

Since the Earth is slowly rotating, the Schwarzschild metric can be used to estimate the frequency difference between atomic clocks on Earth and on satellites, which is important e.g. to calculate local positions from satellite signals.

8.5 Escape velocities and infall times

The study of radial infall and escape trajectories in the Schwarzschild spacetime is interesting in its own right, and also helps us to get more practice and insight into the physical meaning and use of the components of the metric tensor.

Escape velocity in the Schwarzschild spacetime

We have seen that a particle must have $\gamma = E/mc^2 \geqslant 1$ to be able to escape to infinity. Suppose an astronaut is at rest at a certain value of r. With what initial speed v_{esc} (from her point of view) would the astronaut have to toss out a particle in radial direction for the particle to just escape to $r \to \infty$?

We have $J = 0$ (vanishing angular momentum, $\dot{\varphi} = 0$) since the particle is moving in radial direction. Equation (8.29) yields for $\gamma = 1$, $J = 0$ and $dr/d\tau > 0$

$$\frac{dr}{d\tau} = \sqrt{\frac{2GM}{r}} = c\sqrt{\frac{r_S}{r}}. \tag{8.90}$$

However, this does not yet define the physical escape speed of the particle, since the astronaut would measure this speed as

$$v = \frac{d(\text{length in radial direction})}{d(\text{time } \tau' \text{ measured by the astronaut})} = \sqrt{g_{rr}}\frac{dr}{d\tau'}. \tag{8.91}$$

Remember that τ is the eigentime of the particle moving along the radial geodesic:

$$d\tau = \sqrt{\frac{r - r_S}{r}dt^2 - \frac{r}{r - r_S}\frac{dr^2}{c^2}}, \tag{8.92}$$

which implies for the Schwarzschild coordinate time

$$\begin{aligned} dt &= \sqrt{\frac{r}{r - r_S}d\tau^2 + \frac{r^2}{(r - r_S)^2}\frac{dr^2}{c^2}} = d\tau\sqrt{\frac{r}{r - r_S} + \frac{r^2}{c^2(r - r_S)^2}\frac{dr^2}{d\tau^2}} \\ &= d\tau\sqrt{\frac{r}{r - r_S} + \frac{rr_S}{(r - r_S)^2}} = d\tau\frac{r}{r - r_S}. \end{aligned} \tag{8.93}$$

The astronaut is at rest at r, and therefore *their eigentime* relates to the Schwarzschild coordinate time through

$$d\tau' = dt\sqrt{\frac{r - r_S}{r}} = d\tau\sqrt{\frac{r}{r - r_S}}. \tag{8.94}$$

This yields for the escape speed measured by the astronaut

$$v_{esc} = \sqrt{g_{rr}}\frac{dr}{d\tau'} = \sqrt{g_{rr}\frac{r - r_S}{r}}\frac{dr}{d\tau} = \sqrt{\frac{2GM}{r}} = c\sqrt{\frac{r_S}{r}}, \tag{8.95}$$

which coincides with the classical result. It also coincides with $dr/d\tau$ (8.90), but only because $g_{rr}g_{00} = -1$ in the Schwarzschild spacetime. We note in particular that escape from the Schwarzschild horizon $r = r_S$ would require $v_{\text{esc}} = c$.

Radial infall time into the Schwarzschild horizon—the case of massive particles

For an infalling astronaut or massive particle with $\gamma = 1$ and $J = 0$ equation (8.29) yields

$$\frac{dr}{d\tau} = -\sqrt{\frac{2GM}{r}} = -c\sqrt{\frac{r_S}{r}}. \tag{8.96}$$

The freely falling astronaut or massive particle covers the distance between some finite radius r_0 and the event horizon r_S after a finite eigentime

$$\tau = \frac{1}{c}\int_{r_S}^{r_0} dr\,\sqrt{\frac{r}{r_S}} = \frac{2}{3c\sqrt{r_S}}\left(\sqrt{r_0}^3 - \sqrt{r_S}^3\right). \tag{8.97}$$

However, for the coordinate time t we find with equation (8.27)

$$\frac{dr}{dt} = -c\frac{r - r_S}{r}\sqrt{\frac{r_S}{r}}, \tag{8.98}$$

corresponding to an infall time from radius r_0 to radius r_1

$$t = \frac{1}{c\sqrt{r_S}} \int_{r_1}^{r_0} dr \frac{r\sqrt{r}}{r - r_S} = \frac{1}{c\sqrt{r_S}} \left[\frac{2}{3}\sqrt{r_0}^3 - \frac{2}{3}\sqrt{r_1}^3 + 2r_S\sqrt{r_0} - 2r_S\sqrt{r_1} + \sqrt{r_S}^3 \ln\left(\frac{\sqrt{r_0} - \sqrt{r_S}}{\sqrt{r_1} - \sqrt{r_S}} \frac{\sqrt{r_1} + \sqrt{r_S}}{\sqrt{r_0} + \sqrt{r_S}} \right) \right]. \tag{8.99}$$

The Schwarzschild time t (8.99) for infall diverges logarithmically for $r_1 \to r_S$, see also figure 8.2.

An infalling observer or particle reaches the Schwarzschild radius after a finite amount of eigentime has elapsed, and the derivatives of the equations (8.26) and (8.29) show that the infalling object does not experience diverging accelerations at the Schwarzschild horizon. It can also be shown that none of the curvature invariants diverges at $r = r_S$, and therefore the Schwarzschild horizon is penetrable for an observer. However, inside the horizon the effective potentials for radial motion that we have found were always strongly attractive $\sim -r^{-3}$ while at the same time r becomes a timelike coordinate. We therefore conclude that inside the horizon everything is drawn forward in the $(-r)$ time direction into the singularity $r = 0$. Since nothing can move away from the singularity, nothing can leave the region inside the Schwarzschild radius, and this explains why the Schwarzschild radius is denoted as an *event horizon*. Note also that inside the horizon the Schwarzschild metric is not static any more (albeit still t-independent), since $-r$ is the time variable there.

It is known that burnt out very massive stars with remnant iron core masses $M \gtrsim 3M_\odot$ will collapse without any known mechanism to stop the collapse and

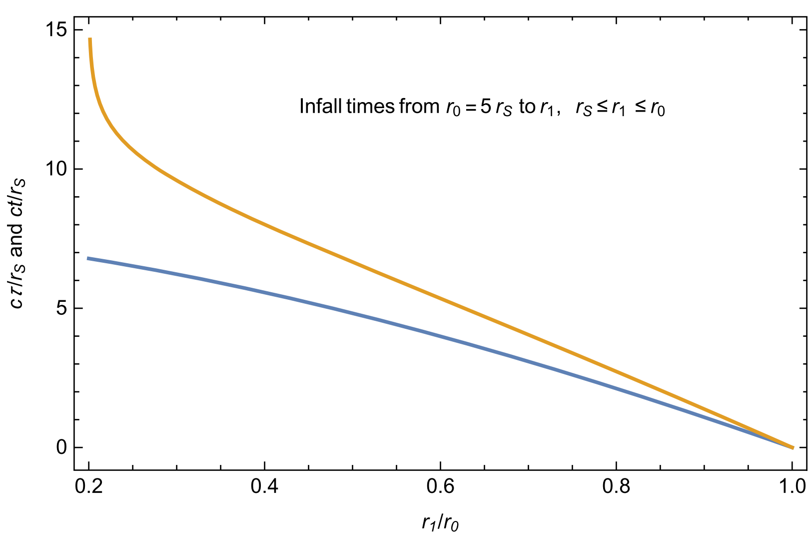

Figure 8.2. The radial infall time from radius $r_0 = 5r_S$ to r_1, $r_S < r_1 \leqslant r_0$, as a function of r_1/r_0. The blue curve is the infall time $c\tau/r_S$ observed on a clock falling towards the horizon, the brown curve is the Schwarzschild infall time ct/r_S observed by a clock at large distance from the black hole.

therefore these stars will end up as black holes. Yet the question whether a particular event horizon exists 'now' depends on the observer's notion of simultaneity. Consider first observers who observe a collapsing star at constant location $\vartheta = \text{const.}$, $\varphi = \text{const.}$ and $r_O = \text{const.}$ with $r_O > r_S$. The eigentime of these observers is related to the Schwarzschild time t through

$$\tau_O = \frac{r_O - r_S}{r_O} t. \tag{8.100}$$

These observers will actually never see the collapsing matter from the star cross into the event horizon at finite time τ_O on their clocks. They will only observe the surface of the star getting ever closer to its Schwarzschild radius and eventually fading out of sight due to ever increasing redshift of the photons emitted by the collapsing matter. For these observers the black hole is like a fading 'frozen star'. However, for all purposes of the observational consequences of the local gravitational field (a.k.a. the local spacetime geometry), the 'frozen star' behaves like the black hole which it approximates in the limit $\tau_O \to \infty$.

On the other hand, the eigentime τ_I of observers who are moving in radial direction, $r = r(\tau_I)$, $\dot{r}(\tau_I) \neq 0$, is related to the Schwarzschild time according to

$$c^2 d\tau_I^2 = \frac{r(\tau_I) - r_S}{r(\tau_I)} c^2 dt^2 - \frac{r(\tau_I)}{r(\tau_I) - r_S} \frac{d^2 r(\tau_I)}{d\tau_I^2} d\tau_I^2, \tag{8.101}$$

and therefore

$$dt = \sqrt{\frac{r(\tau_I) - r_S}{r(\tau_I)} + \frac{1}{c^2} \frac{d^2 r(\tau_I)}{d\tau_I^2}} \frac{r(\tau_I)}{r(\tau_I) - r_S} d\tau_I. \tag{8.102}$$

This can slice the Schwarzschild spacetime in spatial hypersurfaces $\tau_I = \text{const.}$ which intersect the black hole horizon, i.e. for these observers the black hole exists 'now'. Since radial motion relative to the black hole is the generic case for an observer (certainly with respect to most black holes anyway), black holes do exist 'now' for every observer.

Radial infall for arbitrary γ

Equation (8.29) yields for a radially infalling ($J = 0$) astronaut or massive particle the equation

$$\frac{dr}{d\tau} = -c\sqrt{\frac{(\gamma^2 - 1)r + r_S}{r}}, \tag{8.103}$$

and the following possibilities for the infall eigentime from r_0 to r_S:

1. If $\gamma > 1$, i.e. the falling object could start with speed $v_\infty = c\sqrt{\gamma^2 - 1}/\gamma$ at $r \to \infty$:

$$\tau = \frac{1}{c}\int_{r_S}^{r_0} dr \sqrt{\frac{r}{(\gamma^2 - 1)r + r_S}}$$
$$= \frac{1}{(\gamma^2 - 1)c}\left[\sqrt{(\gamma^2 - 1)r^2 + rr_S}\right.$$
$$\left. - \frac{r_S}{\sqrt{\gamma^2 - 1}}\ln\left(\sqrt{(\gamma^2 - 1)r} + \sqrt{(\gamma^2 - 1)r + r_S}\right)\right]_{r_S}^{r_0}. \quad (8.104)$$

2. If $\gamma < 1$:

$$\tau = \frac{1}{(1 - \gamma^2)c}\left[\frac{r_S}{\sqrt{1 - \gamma^2}}\arcsin\left(\frac{2(\gamma^2 - 1)r + r_S}{r_S}\right)\right.$$
$$\left. - \sqrt{(\gamma^2 - 1)r^2 + rr_S}\right]_{r_S}^{r_0}. \quad (8.105)$$

In the latter case r_0 is restricted by $r_0 \leqslant r_S/(1 - \gamma^2)$, but in either case the eigentime for infall from finite radius r_0 is still finite.

For the calculation of the infall time *as measured by a distant observer at rest*, we can convert equation (8.103) through multiplication with $d\tau/dt = (r - r_S)/\gamma r$:

$$\frac{dr}{dt} = -c\frac{r - r_S}{\gamma r}\sqrt{\frac{(\gamma^2 - 1)r + r_S}{r}}. \quad (8.106)$$

This yields for the infall time from r_0 to r_1 with $r_0 > r_1 > r_S$

$$t = \frac{\gamma}{c}\int_{r_1}^{r_0} dr \frac{r\sqrt{r}}{(r - r_S)\sqrt{(\gamma^2 - 1)r + r_S}}. \quad (8.107)$$

This can still be integrated in closed form. However, the important feature for us is that there is always the term $(r - r_S)^{-1}$ in the integrand. Therefore the logarithmic divergence for $r_1 \to r_S$ persists for arbitrary value of γ.

We also cannot avoid this verdict on diverging t by pushing matter together through external forces instead of gravitational collapse: Suppose we force matter into the Schwarzschild horizon on radial trajectories $r = r(\tau)$. Then we still have

$$\frac{dr}{dt} = \frac{d\tau}{dt}\frac{dr}{d\tau} \quad (8.108)$$

and

$$dt = \sqrt{\frac{r}{r - r_S}d\tau^2 + \frac{r^2}{c^2(r - r_S)^2}dr^2} = d\tau\sqrt{\frac{r}{r - r_S} + \frac{r^2}{c^2(r - r_S)^2}\left(\frac{dr}{d\tau}\right)^2}. \quad (8.109)$$

This yields

$$\frac{dr}{dt} = \frac{r - r_S}{r}\left(\sqrt{\frac{r - r_S}{r} + \frac{1}{c^2}\left(\frac{dr}{d\tau}\right)^2}\right)^{-1}\frac{dr}{d\tau}. \tag{8.110}$$

The term $r - r_S$ implies that the logarithmic divergence of the Schwarzschild time t for $r \to r_S$ will prevail.

Chapter 9

Massless particles in the Schwarzschild spacetime

For the discussion of the motion of massless particles we have to overcome the logical difficulty that the eigentime of a massless particle vanishes,

$$c^2d\tau^2 = -ds^2 = -g_{\mu\nu}dx^\mu dx^\nu = 0. \tag{9.1}$$

Due to $d\tau = 0$, *a priori* we also do not have an eigenvelocity $u^\mu = dx^\mu/d\tau$, and therefore we also should not rely on the reasoning with Killing vectors and conservation laws $\epsilon_\mu u^\mu = \text{const.}$ in symmetric spaces to give us first-order equations of motion. This would also not be appropriate since the proof of the conservation law $\epsilon_\mu u^\mu = \text{const.}$ required the second-order equation of motion for x^μ in addition to the Killing equation for ϵ_μ, and it anyhow won't help us if the spacetime has not enough symmetries.

One might hope to circumvent these difficulties by just using a general timelike parameter σ in the action principle

$$\delta S = \delta \int d\sigma \sqrt{-g_{\mu\nu}\frac{dx^\mu}{d\sigma}\frac{dx^\nu}{d\sigma}} = 0. \tag{9.2}$$

However, even then one has to divide by $\sqrt{-g_{\mu\nu}(dx^\mu/d\sigma)(dx^\nu/d\sigma)}$ in the derivation of the equations of motion, and this always vanishes on the actual orbit of a massless particle, $\sqrt{-g_{\mu\nu}(dx^\mu/d\sigma)(dx^\nu/d\sigma)} = cd\tau/d\sigma = 0$.

To avoid all these difficulties we will write the equations for massive particles in a form which is independent both of the mass and the eigentime of the massive particles, and then make the reasonable assumption that these equations should also hold in the massless limit.

At the end of the day this reasoning may seem too clumsy since it finally leads to the conclusion that a class of *affine parameters* τ exists, such that the equation of

doi:10.1088/2053-2571/aaf173ch9

motion for photons has the same form (6.16) as for massive particles (and textbooks discussing lightlike geodesics usually simply start from this assumption). However, besides demonstrating that affine parameters really exist, the present reasoning will also tell us how to actually construct affine parameters.

9.1 Equations of motion

Before we discuss the massless case, recall that from the geodesic equation for a massive particle in terms of its eigentime, we found the following second-order equations (see equations (8.16)–(8.18))

$$\frac{d^2r}{d\tau^2} + \frac{A'}{2A}\left(\frac{dr}{d\tau}\right)^2 - \frac{r}{A}\left(\frac{d\varphi}{d\tau}\right)^2 + \frac{c^2B'}{2A}\left(\frac{dt}{d\tau}\right)^2 = 0, \tag{9.3}$$

$$\frac{d^2\varphi}{d\tau^2} + \frac{2}{r}\frac{d\varphi}{d\tau}\frac{dr}{d\tau} = \frac{1}{r^2}\frac{d}{d\tau}(r^2\dot{\varphi}) = 0, \tag{9.4}$$

$$\frac{d^2t}{d\tau^2} + \frac{B'}{B}\frac{dr}{d\tau}\frac{dt}{d\tau} = \frac{1}{B}\frac{d}{d\tau}(B\dot{t}) = 0. \tag{9.5}$$

We can use equation (9.5), $\partial_\tau \propto (1/B)\partial_t$, to eliminate the eigentime τ of the massive particle. This yields second-order equations for the t dependence of the trajectory:

$$\frac{d^2r}{dt^2} + \left(\frac{A'}{2A} - \frac{B'}{B}\right)\left(\frac{dr}{dt}\right)^2 - \frac{r}{A}\left(\frac{d\varphi}{dt}\right)^2 + \frac{c^2B'}{2A} = 0 \tag{9.6}$$

$$\frac{d}{dt}\left(\frac{r^2}{B(r)}\frac{d\varphi}{dt}\right) = 0. \tag{9.7}$$

However, these equations depend neither on the mass nor on the eigentime of a massive particle, and therefore these equations should also hold in the massless limit. Therefore we can choose these equations as a starting point for the discussion of the trajectories of massless particles in the Schwarzschild geometry.

Equations (9.6) and (9.7) yield again first integrals:

$$\dot{\varphi}(t) = \frac{JB(r(t))}{r^2(t)}, \tag{9.8}$$

$$\frac{A(r(t))}{2B^2(r(t))}\dot{r}^2(t) + \frac{J^2}{2r^2(t)} - \frac{c^2}{2B(r(t))} = K. \tag{9.9}$$

These two equations and $ds^2 = 0$ yield

$$ds^2 = \left(-B + \frac{A}{c^2}\dot{r}^2 + \frac{r^2}{c^2}\dot{\varphi}^2\right)c^2dt^2 = 2B^2Kdt^2 = 0, \tag{9.10}$$

which fixes the constant K for *massless* particles to

$$K = 0. \tag{9.11}$$

This yields for the radial equation

$$\frac{1}{2}\dot{r}^2(t) + \frac{J^2B^2(r(t))}{2A(r(t))r^2(t)} - \frac{c^2B(r(t))}{2A(r(t))} = 0. \tag{9.12}$$

However, we can again use a parameter τ which behaves like the eigentime in the case of the massive particles, if we *define* it through

$$d\tau = B(r(t))dt = \frac{r(t) - r_S}{r(t)}dt. \tag{9.13}$$

Then we find an equation like (8.29), but this time for $K = 0$:

$$\frac{1}{2}\left(\frac{dr}{d\tau}\right)^2 + \left(\frac{1}{2} - \frac{GM}{c^2r}\right)\frac{J^2}{r^2} = \frac{c^2}{2}. \tag{9.14}$$

The corresponding second-order equations for $x^\mu(\tau)$ would be exactly the same equations as the geodesic equations for the massive particles, such that equation (9.13) defines the so called *affine parameters* for the lightlike geodesics in the Schwarzschild spacetime (τ can be rescaled by an arbitrary positive factor).

In a mechanical analog the radial motion of a massless particle in terms of the parameter τ corresponds to the motion of a non-relativistic particle of energy per mass $c^2/2$ and effective potential

$$\begin{aligned}\tilde{V}(r) &= \left(\frac{1}{2} - \frac{GM}{c^2r}\right)\frac{J^2}{r^2}\\ &= \frac{J^2(r - r_S)}{2r^3},\end{aligned} \tag{9.15}$$

see figure 9.1.

This time there is always a centrifugal barrier if $J^2 > 0$, and the tip is always located at $r = 3r_S/2$. This is therefore a lower limit on the periastron (point of closest approach) of a photon which does not fall into the horizon, and this limit coincides with the limit (8.59) for freely falling massive particles[1], which may be not too surprising given that those massive particles approaching that minimal periastron value must be ultrarelativistic (8.61). Contrary to the massive case, this time there does not exist any stable bound orbit for any choice of the parameter J. There are

[1] However, recall that the limits on closest approaches for bound orbits of massive particles (8.65) and circular orbits of massive particles (8.66) are larger.

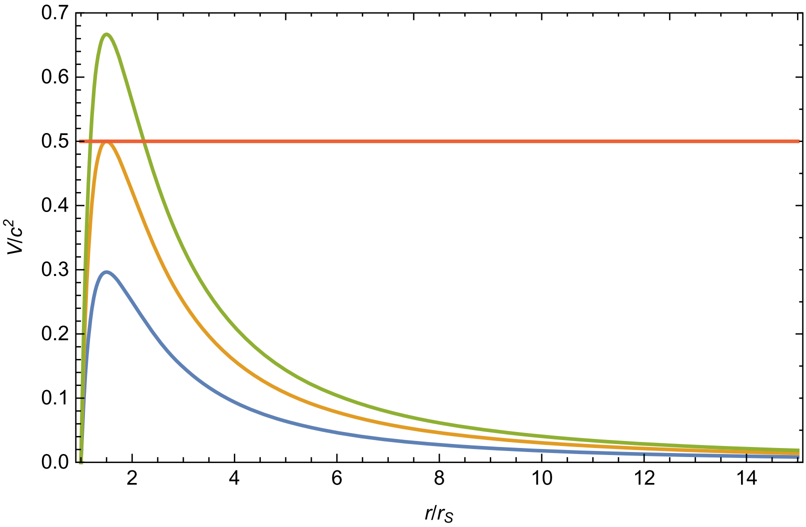

Figure 9.1. The effective potential $\tilde{V}/c^2$ for $r_S \leqslant r \leqslant 15r_S$ and $j = 2$ (blue), $j = \sqrt{27}/2$ (brown), $j = 3$ (green). The value 0.5 for the specific energy in the analog mechanical problem is marked in red.

only unstable circular orbits for photons at $3r_S/2 = 3GM/c^2$ if $J^2 = (27/4)c^2 r_S^2$. This value of J^2 corresponds to the red line in figure 9.1.

The height of the centrifugal barrier is

$$\tilde{V}(3r_S/2) = \frac{2}{27}\frac{J^2}{r_S^2}, \tag{9.16}$$

i.e. infalling photons do not fall into the Schwarzschild horizon if and only if

$$J \geqslant \frac{3\sqrt{3}}{2}cr_S = 3\sqrt{3}\frac{GM}{c}. \tag{9.17}$$

In the massive case $J/v_\infty = b$ is the impact parameter of a particle. This also holds in the present case in the form

$$\frac{J}{c} = \frac{r^2(\tau)\dot{\varphi}(\tau)}{c} = \frac{r^2(t)\dot{\varphi}(t)}{c}\bigg|_{r\to\infty} = \frac{\boldsymbol{r}\times\boldsymbol{v}}{c}\bigg|_{r\to\infty} = b. \tag{9.18}$$

We therefore find that infalling photons do not fall into the Schwarzschild horizon if and only if the impact parameter satisfies

$$b \geqslant \frac{3\sqrt{3}}{2}r_S = 3\sqrt{3}\frac{GM}{c^2}. \tag{9.19}$$

Stated differently, black holes have a photon absorption cross section

$$\sigma = 27\pi r_S^2/4. \tag{9.20}$$

This is legitimate, because b is a perpendicular length parameter at $r \to \infty$ where the Schwarzschild spacetime is asymptotically flat, and therefore b is a *bona fide* length perpendicular to the infalling photon path. It is not a radial value near the black hole, where g_{rr} only allows us define radial distances from the Schwarzschild horizon.

9.2 Deflection of light in a gravitational field

We now consider photons which escape again to infinity, i.e. the impact parameter satisfies $b > \frac{3\sqrt{3}}{2} r_S$. In the massless case, equation (9.14) and $\dot{\varphi} = J/r^2$ imply

$$\left(\frac{dr}{d\varphi}\right)^2 - \frac{c^2}{J^2} r^4 + r^2 - r_S r = 0. \tag{9.21}$$

This yields again an elliptic integral,

$$\begin{aligned} \varphi - \varphi_0 &= \pm \int_{r(\varphi_0)}^{r(\varphi)} dr \frac{J}{\sqrt{r(c^2 r^3 - J^2 r + J^2 r_S)}} \\ &= \pm \int_{r(\varphi_0)}^{r(\varphi)} dr \frac{b}{\sqrt{r(r^3 - b^2 r + b^2 r_S)}}. \end{aligned} \tag{9.22}$$

Now we assume that in the (x, y)-plane with the mass M in the center, the initial photon infall is from $x \to -\infty$ along the line $y = -$b. This implies that the initial angle is $\varphi_0 = -\pi$, $r(\varphi_0) \to \infty$, and r decreases with increasing φ (i.e. we have to take the negative sign in equation (9.22)) until it reaches the minimal value r_1 with corresponding angle

$$\varphi(r_1) = -\pi - \int_{\infty}^{r_1} dr \frac{b}{\sqrt{r(r^3 - b^2 r + b^2 r_S)}}. \tag{9.23}$$

The deflection angle of the photon is then

$$\begin{aligned} \Delta\varphi = \varphi(r \to \infty) &= \varphi(r_1) + \int_{r_1}^{\infty} dr \frac{b}{\sqrt{r(r^3 - b^2 r + b^2 r_S)}} \\ &= -\pi - 2\int_{\infty}^{r_1} dr \frac{b}{\sqrt{r(r^3 - b^2 r + b^2 r_S)}}. \end{aligned} \tag{9.24}$$

To calculate the deflection angle for given impact parameter b from this equation requires knowledge of the radial coordinate r_1 of the point of closest approximation. The condition $dr/d\varphi|_{r=r_1} = 0$ for the point of closest approximation yields

$$(r_1 - r_S)b^2 = r_1^3. \tag{9.25}$$

We solve this equation in first order under the assumption $r_S \ll b$. The zeroth-order solution is $r_1|_{r_S=0} = b$, and the first-order *Ansatz* $r_1 = b + \epsilon$ yields $\epsilon = -r_S / 2$:

$$r_1 \approx b - \frac{1}{2} r_S. \tag{9.26}$$

The zeroth-order term in the deflection angle vanishes,

$$\Delta\varphi|_{r_S=0} = -\pi - 2\int_\infty^b dr \frac{b}{r\sqrt{r^2 - b^2}} = -\pi + 2 \arccos \frac{b}{r}\Big|_\infty^b = 0, \tag{9.27}$$

as it should, since $r_S = 0$ means no deflecting mass. Here the parameter range is $-\pi \leqslant \varphi = \arccos(b/r) \leqslant 0, \arccos 0 = -\pi/2$. This leaves us with the first-order term[2]

$$\begin{aligned}\Delta\varphi &\simeq \lim_{r\to b} \frac{r_S}{\sqrt{r^2 - b^2}} + \int_\infty^b dr \frac{b^3 r_S}{r^2 \sqrt{r^2 - b^2}^3} \\ &= \lim_{r\to b} \frac{r_S}{\sqrt{r^2 - b^2}} - r_S \frac{2r^2 - b^2}{br\sqrt{r^2 - b^2}}\Big|_\infty^b \\ &= \frac{2r_S}{b},\end{aligned} \tag{9.29}$$

i.e. the deflection angle of a photon of impact parameter b in the gravitational field of a slowly rotating mass M is

$$\Delta\varphi \simeq \frac{2r_S}{b} = \frac{4GM}{c^2 b}. \tag{9.30}$$

This effect is often used to measure the masses of galaxies and galaxy clusters through gravitational lensing.

Gravitational lensing

Suppose we observe light from a distant bright astrophysical object (typically a quasar), and a galaxy or galaxy cluster of mass M is close to the path of the light from the distant quasar to us. On its way to us the light from the distant quasar will be deflected at the mass M by a deflection angle (9.30)

$$\alpha \simeq \frac{4GM}{c^2 b} = \frac{2r_S}{b} \simeq \frac{2r_S}{\Theta d_{MO}}. \tag{9.31}$$

Here Θ is the small angle between the incident light rays from the quasar and the line of sight to the intervening mass M, and d_{MO} is the distance between us and the

[2] Note that

$$\int^{f(x+\epsilon)} d\xi\, I(\xi) \simeq \int^{f(x)} d\xi\, I(\xi) + \epsilon f'(x) I(f(x)). \tag{9.28}$$

The second term in equation (9.29) comes from the expansion of the integrand in equation (9.24) with respect to r_S (or rather $r_S/b \ll 1$).

mass M. The impact parameter $b \simeq d_{MO}\Theta$ is the distance between the gravitational lens M and the passing light ray. The geometric setup is displayed in figure 9.2.

If β is the angle between the line of sight to M and the undistorted line of sight to the distant quasar, we have for small angles

$$\Theta(d_{MO} + d_{SM}) = \beta(d_{MO} + d_{SM}) + \alpha d_{SM}, \tag{9.32}$$

where d_{SM} is the distance between the quasar and the mass M. Substitution of α from equation (9.31) yields the gravitational *lens equation*:

$$\Theta = \beta + \frac{2r_S d_{SM}}{\Theta d_{MO}(d_{MO} + d_{SM})} = \beta + \frac{\Theta_E^2}{\Theta}. \tag{9.33}$$

The *Einstein angle*

$$\Theta_E = \sqrt{\frac{2r_S d_{SM}}{d_{MO}(d_{MO} + d_{SM})}} \tag{9.34}$$

is the deflection angle if the source is located directly behind the mass M, and in the ideal case can be observed as the opening angle of an *Einstein ring* around the mass M. The ring is the picture of the quasar.

In the generic case the two solutions of the lens equation (9.33) are

$$\Theta_\pm = \frac{\beta}{2} \pm \sqrt{\frac{\beta^2}{4} + \Theta_E^2}, \tag{9.35}$$

and we expect to see two images: one at $\Theta_+ > 0$ on the same side as the source, and one at $\Theta_- < 0$ on the opposite side, as shown in figure 9.3.

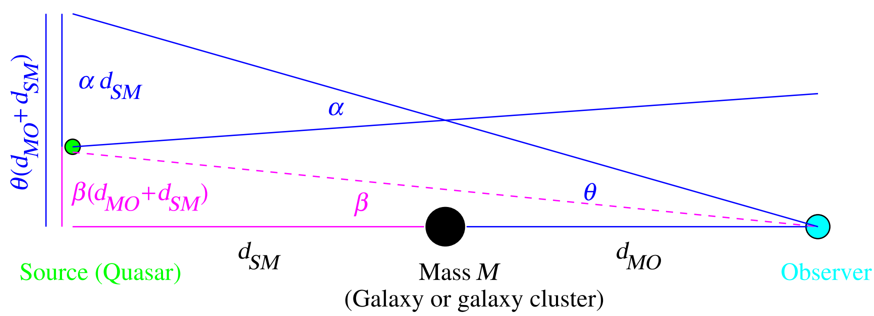

Figure 9.2. The source–lens–observer system for a gravitational lens.

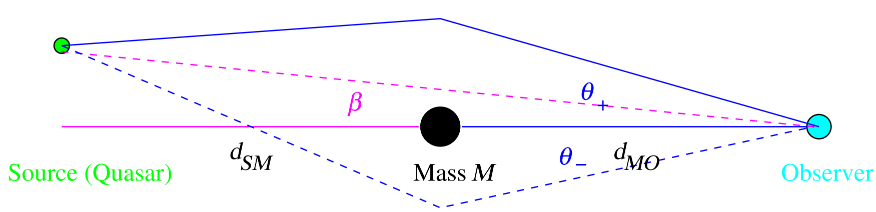

Figure 9.3. Two pictures of a distant quasar from a gravitational lens with mass M.

The identification of two (or more) pictures of an astrophysical object as really arising from one and the same source can involve reasoning based on

- ringlike deformations of pictures around the lensing mass as an indication of a lensed picture,
- same redshift and spectral features in the two pictures,
- eventually a correspondence of local intensity distributions across the two pictures,
- eventually similarity in intensity fluctuations in the two pictures (where one has to take into account that the difference in light paths will lead to a time delay between the two pictures).

If two pictures have been identified to come from the same source, they can easily be assigned to the two angles Θ_+ and Θ_-due to $|\Theta_-| < \Theta_+$, i.e. the picture on the opposite side will be closer to the line of sight to the lensing mass M.

The two angles then determine the real line of sight to the distant object and the mass $M = c^2 r_S/2G$ of the lens through

$$\Theta_+ + \Theta_- = \Theta_+ - |\Theta_-| = \beta, \tag{9.36}$$

and

$$-\Theta_+\Theta_- = \Theta_+|\Theta_-| = \Theta_E^2 = \frac{2r_S d_{SM}}{d_{MO}(d_{MO} + d_{SM})}. \tag{9.37}$$

The determination of M from the last equation assumes that the distances d_{MO} and $d_{MO} + d_{SM}$ to the lens and the source have been determined through appropriate distance indicators[3].

The prediction of two pictures assumes spherical symmetry of the mass distribution in the lens, and that the deflecting mass M is contained within the impact parameters of the two light rays above and below M. Often only one picture or more than two pictures can be observed, due to deviation of M from spherical shape or because the deflecting mass distribution extends well into the light path.

Eddington[4] pointed out in 1920 that astrophysical objects could generate more than one picture in a telescope due to light deflection at a large mass close to the line of sight. The first gravitational lens was discovered in 1979 by Walsh, Carswell and Weymann[5].

[3] Astronomers have several tools to determine the distances to astronomical objects. In the billions of light-years range this would include in particular the use of type 1a supernovae as standard candles, or the use of the cosmological distance-redshift relation. Tools in the million to billion light-years range would also include the use of Cepheid variable stars as standard candles, maser astrometry, and the use of the Tully–Fisher relation for rotating galaxies.

[4] Eddington A S 1920 *Space, Time and Gravitation* (Cambridge: Cambridge University Press). See also Renn J, Sauer T and Stachel J 1997 *Science* **275** 184.

[5] Walsh D, Carswell R F and Weymann R J 1979 *Nature* 279 381.

9.3 Apparent photon speeds and radial infall

Since clocks and rulers depend on the position and motion in a gravitational field, it is interesting to see how different observers judge the motion of photons. We will therefore consider local light speeds and photon infall time in the Schwarzschild spacetime.

Apparent photon speeds in the Schwarzschild spacetime

From equations (9.8) and (9.12), or from $ds^2 = 0$, we find for the apparent speed of the photon at radius $r > r_S$ relative to a *distant* observer at rest ($r_{\text{Observer}} \to \infty$) the expression

$$v^2 = A(r)\left(\frac{dr}{dt}\right)^2 + r^2\left(\frac{d\varphi}{dt}\right)^2 = c^2 B(r) = c^2 - 2\frac{GM}{r} = c^2\frac{r - r_S}{r}. \tag{9.38}$$

This corresponds to an effective local refraction index

$$n = \frac{c}{v} = \sqrt{\frac{r}{r - r_S}} > 1, \tag{9.39}$$

and the gravitational field looks like an optically dense medium for the observer.

An observer at rest at *finite* radius $r_O > r_S$ measures time intervals

$$d\tau'^2 = \frac{r_O - r_S}{r_O}dt^2, \tag{9.40}$$

and therefore they measure an apparent speed for the photon at radius r which is given by

$$v'^2 = \left(\frac{dt}{d\tau'}\right)^2 v^2 = \frac{r_O}{r_O - r_S}v^2 = c^2\frac{r - r_S}{r_O - r_S}\frac{r_O}{r}. \tag{9.41}$$

In particular they again measure speed c for all photons close to them. According to that observer's time standards, photons moving closer to r_S are moving slower than c while those moving at higher altitude, $r > r_O$, move faster than c. However, they still agree with all other observers that nothing can move faster than photons in the sense that nothing can overtake the photons at any radius r.

We can also make the following general observations without restriction to the Schwarzschild spacetime:

Assume an observer who uses comoving coordinates in the sense that they are at rest in their own coordinate system. The metric in their coordinate system is

$$ds^2 = g_{\mu\nu}dx^\mu dx^\nu = g_{00}c^2dt^2 + 2g_{0i}cdtdx^i + g_{ij}dx^i dx^j, \tag{9.42}$$

and their eigentime is given by

$$d\tau^2 = -\frac{ds^2}{c^2}\bigg|_{dx^i=0} = -g_{00}\,\big|_{\text{Observer}}\, dt^2, \tag{9.43}$$

where the label 'Observer' denotes the metric component at their location. $ds^2 = 0$ yields the speed of photons measured by them,

$$\begin{aligned} v^2 = g_{ij}(x)\frac{dx^i}{d\tau}\frac{dx^j}{d\tau} &= -\frac{g_{ij}(x)}{g_{00}\,|_{\text{Observer}}}\frac{dx^i}{dt}\frac{dx^j}{dt} \\ &= \frac{g_{00}(x)}{g_{00}\,|_{\text{Observer}}}c^2 + 2\frac{g_{0i}(x)}{g_{00}\,|_{\text{Observer}}}c\frac{dx^i}{dt}. \end{aligned} \tag{9.44}$$

Here x is the location of the photon when the observer measures its speed. In particular, if they use a coordinate system with $g_{0\text{i}} = 0$ in their neighborhood (e.g. a local inertial frame), then they measure $v = c$ for all the photons close to them, because $g_{00}|_{\text{Observer}} \simeq g_{00}(x)$ in her neighbourhood.

Radial infall time into the Schwarzschild horizon of a black hole—the case of massless particles

Equation (9.12) yields the radial speed as a function of Schwarzschild time t,

$$\frac{1}{2}\left(\frac{dr}{dt}\right)^2 + \frac{1}{2}\left(1 - \frac{r_S}{r}\right)^3\frac{J^2}{r^2} - \frac{c^2}{2}\left(1 - \frac{r_S}{r}\right)^2 = 0. \tag{9.45}$$

Recall that the affine time parameter τ (9.13) is *not* an eigentime for photons (as photons have no eigentime). Nevertheless, we notice that the radial infall time parameter τ for photons from $r = r_0 > r_S$ to $r = r_S$, which follows from equation (9.14) for $J = 0$, is

$$\frac{dr}{d\tau} = -c \quad \Rightarrow \quad \tau = \frac{r_0 - r_S}{c}. \tag{9.46}$$

On the other hand, the coordinate time t (measured by a stationary clock at $r \to \infty$) which elapses while the photon is radially falling from $r_0 > r_S$ to a smaller radius $r_1 > r_S$ follows from equation (9.45) with $J = 0$,

$$\frac{dr}{dt} = -c\frac{r - r_S}{r}, \tag{9.47}$$

as

$$ct = -\int_{r_0}^{r_1} dr\frac{r}{r - r_S} = r_0 - r_1 + r_S \ln\left(\frac{r_0 - r_S}{r_1 - r_S}\right). \tag{9.48}$$

This is the same result as for a massive particle (8.107) in the limit $\gamma \to \infty$, and we see again the logarithmic divergence of the radial infall time as measured by a stationary clock.