General Relativity
version 2024

	email address:	Martin.Kunz@unige.ch
Martin Kunz	office:	Ecole de Physique 224
	phone number:	+41 (0)22 379 6350

August 29, 2025

Contents

Preface

Throughout these lecture notes we will use the metric signature $\{-,+,+,+\}$.

I am very grateful to Ruth Durrer, Pierre Fleury and Peter Wittwer for helpful discussions and comments, as well as to Edwin Genoud-Prachex, Ahmadreza Nourizonoz, Julien Songeon and Dorian Wenzel for pointing out typos in the manuscript. I also thank my assistants Joyce Byun, Chris Finlay, Mona Jalilvand, Francesca Lepori, Tobias Nadolny, Ahmadreza Nourizonoz, Viraj Nistane and especially Pierre Fleury for their help with the course during these years, and for preparing and teaching all the exercise sessions.

Chapter 1

Constructing General Relativity

1.1 Introduction

1.1.1 Space and time

For most of us the pre-relativity notion of space and time as a fixed stage on which events happen, and where the history of the world is played out, appears completely natural and obvious. We have no problem to imagine the entire Universe evolving in such a setting, and to imagine simultaneous events happening light-years apart. However, as mankind figured out around the end of the 19th and the beginning of the 20th century, this is wrong. In reality there is no notion of universal simultaneity, as Einstein explained in the theory of Special Relativity.

In the pre-relativity world we can distinguish three classes of pairs of events A and B (with non-zero spatial separation),

1. Event A is in the future of event B. In principle an agent can be present at B and then later at A.
2. Event A is in the past of B. In principle an agent can be present at A and then later at B.
3. Events A and B are simultaneous, it is impossible to be present at A and B for a single agent.

It is useful to introduce the notion of *inertial observers*. These are observers that are not subject to any external forces and thus move "passively" along space-time. Inertial observers can construct a coordinate system for themselves by carrying a rigid, non-rotating three-dimensional ("x, y, z") grid with synchronized clocks, which then allows to label events in space-time with coordinates $x^\mu = (t, x, y, z)$. These labels are called global inertial coordinates.

In pre-relativistic physics, if two such inertial observers O and O' (moving with a relative velocity v in the x direction) compare their coordinate systems (normalized so that they meet at the event $t = 0, x = 0, y = 0, z = 0$ in both coordinate systems), they would find that their coordinates are related by a Galilean transformation,

$$t' = t\,, \tag{1.1}$$
$$x' = x - vt\,, \tag{1.2}$$
$$y' = y\,, \tag{1.3}$$
$$z' = z\,. \tag{1.4}$$

We can see that the sets of simultaneous events for observer O, namely those on the same surface $t = \mathrm{constant}$, are the same as those for observer O'.

However, Special Relativity (where the clock synchronisation is a bit less trivial) says that the two coordinate systems are actually related by a Lorentz transformation,

$$t' = (t - vx/c^2)/\sqrt{1 - v^2/c^2}\,, \tag{1.5}$$
$$x' = (x - vt)/\sqrt{1 - v^2/c^2}\,, \tag{1.6}$$
$$y' = y\,, \tag{1.7}$$
$$z' = z\,. \tag{1.8}$$

Here c is the speed of light, and we will usually chose units so that $c = 1$. In this case the hypersurfaces of $t = \text{constant}$ and of $t' = \text{constant}$ do not coincide, i.e. the two observers will consider different events as simultaneous.

There is no reason to prefer the coordinates of any inertial observer to those of any other such observer. An important question is thus, what are the space-time quantities that are the same for all observers? For pre-relativistic physics there are two such quantities, the time elapsed between two events, Δt, and the spatial distance, $|\Delta \mathbf{x}|$ between two simultaneous events (where we remind the reader that all "pre-relativistic" observers agree on which events are simultaneous).

In relativistic physics however neither of these quantities are invariant – observers do not even in general agree on which events are simultaneous. Instead it is the space-time distance

$$\Delta s^2 \equiv -(\Delta t)^2 + \frac{1}{c^2}\left((\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2\right) \tag{1.9}$$

that remains invariant (and is indeed the only observer-independent quantity that characterises the space-time relationship between events). The Poincaré transformations are then the linear transformations that leave Δs^2 unchanged. In analogy with the Euclidean metric of the usual flat space, we associate to Δs^2 the metric of space-time in Special Relativity, by writing

$$c^2 \Delta s^2 = \sum_{\mu,\nu=0}^{3} \eta_{\mu\nu} (\Delta x)^\mu (\Delta x)^\nu \tag{1.10}$$

where we set $(\Delta x)^0 = c\Delta t$ and where $\eta_{\mu\nu} = \text{diag}(-1, 1, 1, 1)$.

After the development of Special Relativity, it was necessary to modify the laws of physics to be in agreement with the new concept of space-time. Maxwell's theory of electromagnetism was already consistent with SR, this was after all one of the reasons to develop SR in the first place as electrodynamics was not compatible with the pre-relativistic structure of space-time and notion of invariance under Galilean transformations. Newton's theory of gravity on the other hand relies on the notion of instantaneous influence of one body on another, which is not compatible with a theory like Special Relativity where the notion of simultaneity does not exist independently of the observer. But maybe Newton's gravity could be fit into the framework of SR?

At least two reasons convinced Einstein that this was the wrong approach. We discuss the first one, the notion of universal free fall, in more detail below. The second one, less precise, is called *Mach's principle*. In SR and in pre-relativity physics, the structure of space-time is fixed in advance and is not affected by the contents of the universe. In particular, the concepts of 'inertial motion' and 'non-rotating' do not depend on the matter present. Mach, and other scholars, found this unsatisfactory, Mach thought that the local definition of 'non-accelerating' and 'non-rotating' should depend on the matter in the universe. These ideas motivated Einstein to look for a theory where the structure of space-time is influenced by the presence of matter.

1.1.2 Universality of free fall and the equivalence principle

In principle the inertial mass that appears in Newton's second law of motion, $\mathbf{F} = m\mathbf{a}$, and the gravitational mass that is found in Newton's law of gravitation, $F = GmM/r^2$ could be different. However, experiments since the 16th century (well before Newton) have shown that objects fall at the same rate, independent of their composition (if air resistance is negligible)[1], which implies that the two masses are the same. Normally such a coincidence (experimentally verified to better than $1 : 10^{13}$ today) needs an explanation, but in Newton's theory there is no reason for it. This was one motivation for Einstein to look for a theory that would naturally explain the equality.

The universality of free fall can be encoded in a series of *equivalence principles* of varying strictness. The *Weak Equivalence Principle* (WEP) states that:

> The motion of a test body in a gravitational field is independent of its mass and composition (at least when one neglects interactions of spin or of a quadrupole moment with field gradients).

A stronger version is called *Einstein's Equivalence Principle* (EEP),

> In an arbitrary gravitational field no local non-gravitational experiment can distinguish a freely falling non-rotating system (local inertial system) from a uniformly moving system in the absence of a gravitational field.

In other words, gravity can locally be transformed away. The formulation is here not very precise, we will revisit the equivalence principle later when we construct the theory of General Relativity. Among other things, the EEP implies that gravity and inertia cannot be uniquely separated.

1.1.3 Why we need a new theory

We can apply the EEP to the famous Einstein elevator thought experiment. In this thought experiment we are sitting inside an elevator in a static gravitational field (that we take to be homogeneous for simplicity, with acceleration g). We assume that the elevator starts falling freely at time $t = 0$ and that a photon of frequency ν is emitted at the same time from the ceiling towards the floor. The EEP now implies that the the light arrives at the floor at time $t = h/c$ where h is the height of the elevator cabin, and that there is no frequency shift observed by a (freely falling) observer inside the cabin.

However, consider an external observer at rest in the elevator shaft, at the precise point where the photon arrives at the floor of the elevator. The observer at rest moves with respect to the freely falling observer with velocity $v = gt$. Hence the external observer sees the photon Doppler-shifted towards the blue by (to first order)

$$z \equiv \frac{\Delta\nu}{\nu} \simeq \frac{v}{c} = \frac{gh}{c^2} \,. \tag{1.11}$$

This consequence of the EEP has been tested experimentally, e.g. to a level of better than 10^{-4} by a hydrogen maser clock on a rocket. But also, for example, the GPS system would fail within hours if the resulting gravitational time dilation would not be accounted for correctly.

The blue-shift is also necessary for energy conservation: if there was no such shift, we could let a mass m fall in a gravitational field, convert its total energy (including mass energy) to a photon that is sent back to the top, where it is again converted into a mass that falls, etc. If the photon does not lose energy on its way up then we gain the energy mgh in each cycle, violating energy conservation. The photon redshift

[1] See e.g. the Apollo 15 video for a more modern demonstration.

predicted by the above thought experiment precisely cancels the energy gain:

$$E_{\mathrm{down}} = E_{\mathrm{up}} + mgh = mc^2 + mgh = E_{\mathrm{up}}\left(1 + \frac{gh}{c^2}\right), \tag{1.12}$$

$$1 + z = \frac{\lambda_{\mathrm{up}}}{\lambda_{\mathrm{down}}} = \frac{h\nu_{\mathrm{down}}}{h\nu_{\mathrm{up}}} = \frac{E_{\mathrm{down}}}{E_{\mathrm{up}}} = 1 + \frac{gh}{c^2}. \tag{1.13}$$

Why is this a problem for a theory of gravity built on the space-time of Special Relativity? In Special Relativity, a clock moving along a time-like world line $x^\mu(t)$ (parameterised by an arbitrary parameter t) measures the proper time interval

$$\Delta\tau = \int_{t_1}^{t_2} \sqrt{-\eta_{\mu\nu}\frac{dx^\mu}{dt}\frac{dx^\nu}{dt}}dt \tag{1.14}$$

where $\eta_{\mu\nu}$ is the Minkowski metric mentioned above.

Let us assume that a special relativistic theory of gravity exists. We perform a gravitational redshift experiment in a static field, with emitter and absorber moving on world lines of constant height z. The emitter sends photons at a fixed frequency from time t_1 to t_2. The photons will follow world lines to the receiver that are not necessarily straight lines at a $45°$ angle due to the presence of the gravitational field, but as the situation is static, the world lines will be parallel. This means that in a flat Minkowski metric and for proper time intervals given by Eq. (1.14) the arrival times t'_1 and t'_2 will have the same time difference as the emission times, $t'_2 - t'_1 = t_2 - t_1$. In other words, there would be no redshift, which would contradict the EEP and energy conservation, or else (1.14) cannot be valid. Clearly the latter has to be the case, Special Relativity is thus not able to serve as the basis for a new theory of gravity.

The way Einstein ended up solving this problem and incorporating elements from Mach's principle was to move towards a dynamic space-time where the metric in (1.14) changes as a function of the gravitational field. In the next section we will set up the mathematical machinery to deal with this situation, before constructing the theory of General Relativity. We will try hard to make the whole development of General Relativity look simple, or at least straightforward. But in reality it was neither. It took Einstein nearly 10 years of exploring ways to build a theory that would work in a satisfactory way, in 1915 he wrote on a postcard to his friend Michele Besso *"I worked horribly strenuously, strange that one can endure that"* (as quoted in [2]).

1.2 Mathematical interlude

A crucial feature of GR will be the curved space-time in which inertial observers follow geodesics. In order to formulate such a theory, we need to introduce the mathematical structure that allows to describe such features. However, we want to focus primarily on GR, for this reason this section is not intended to be a complete, or even very rigorous, introduction to differential geometry. More mathematical and complete treatments can be found for example in [3] and [2].

1.2.1 Manifolds

We can easily imagine a lower dimensional curved space, e.g. the surface of a sphere, that is embedded in our three-dimensional space. However, we do not want to limit ourselves to spaces that can be embedded in a flat higher dimensional space. Instead, it is possible to use manifolds to provide an "intrinsic" description.

A manifold can be constructed by glueing together local descriptions, maps, that cover the complete manifold. A typical example is the way in which we describe the surface of the Earth with the help of local maps that can be approximated as being flat. To be more precise, an n-dimensional C^∞ real manifold $\mathcal{M}$ is a set of points, together with a collection of subsets $\{\mathcal{O}_\alpha\}$ that satisfy the following properties:

1. Each point $p \in \mathcal{M}$ lies in at least one subset $\mathcal{O}_\alpha$, i.e. the $\{\mathcal{O}_\alpha\}$ cover $\mathcal{M}$.
2. For each α there is a one-to-one, onto, map $\psi_\alpha : \mathcal{O}_\alpha \to U_\alpha$ where U_α is an open subset of $\mathbb{R}^n$ (these are the maps).
3. If any two subsets $\mathcal{O}_\alpha$ and $\mathcal{O}_\beta$ overlap so that their intersection is not the empty set, we can consider the map $\psi_\beta \circ \psi_\alpha^{-1}$ that maps points in the intersection from U_α to U_β, i.e. it is an application from $\mathbb{R}^n$ to $\mathbb{R}^n$. We require this map to be C^∞.

The applications ψ_α that generate the maps of $\mathcal{M}$ are called *charts* by mathematicians and *coordinate systems* by physicists. One often requires that the cover $\{\mathcal{O}_\alpha\}$ and the associated charts ψ_α are maximal, i.e. all coordinate systems compatible with the definition above are included, in order to make the definition of a manifold more unique.[2]

A trivial example of a manifold is $\mathbb{R}^n$ which can be covered with a single chart by using $\mathcal{O} = \mathbb{R}^n$ and $\psi =$ identity map. The standard non-trivial example is the two-sphere S^2,

$$S^2 = \{(x^1, x^2, x^3) \in \mathbb{R}^3 | (x^1)^2 + (x^2)^2 + (x^3)^2 = 1\} . \tag{1.15}$$

S^2 cannot be covered with a single chart like $\mathbb{R}^n$, but we can e.g. choose six hemispherical open sets,

$$\mathcal{O}_i^\pm = \{(x^1, x^2, x^3) \in S^2 | \pm x^i > 0\} . \tag{1.16}$$

The set of the $\mathcal{O}_i^\pm$ obviously covers S^2. We can choose various ways to define charts $\mathcal{O}_i^\pm \to \mathbb{R}^2$, for example simply $f_1^+ : \mathcal{O}_1^+ \to \mathcal{D}$; $f_1^+(x^1, x^2, x^3) = (x^2, x^3)$, and so on, where $\mathcal{D}$ is the open disk, $\mathcal{D} = \{(y^1, y^2) \in \mathbb{R}^2 | (y^1)^2 + (y^2)^2 < 1\}$. Within each hemispherical set one can uniquely reconstruct all three coordinates, and using the explicit expressions it is straightforward to show that the overlap functions that map points from one hemispherical set to another one are differentiable, for example:

$$\begin{aligned} &\left(f_1^+\right)^{-1} : (x^2, x^3) \to (\sqrt{1 - (x^2)^2 - (x^3)^2}, x^2, x^3) \\ &\Rightarrow f_3^+ \circ \left(f_1^+\right)^{-1} : (x^2, x^3) \to (\sqrt{1 - (x^2)^2 - (x^3)^2}, x^2) . \end{aligned} \tag{1.17}$$

We can also define maps between manifolds: Let $\mathcal{M}$ and $\mathcal{M}'$ be manifolds, with chart maps ψ_α and ψ'_β. A map $f : \mathcal{M} \to \mathcal{M}'$ is then called C^∞ if for each α and β the map $\psi'_\beta \circ f \circ \psi_\alpha^{-1}$ from $U_\alpha \subset \mathbb{R}^n$ to $U'_\beta \subset \mathbb{R}^{n'}$ is C^∞ in the usual sense for applications $\mathbb{R}^n \to \mathbb{R}^{n'}$. If $f : \mathcal{M} \to \mathcal{M}'$ is C^∞, one-to-one, onto and has a C^∞ inverse, then f is called a diffeomorphism, and $\mathcal{M}$ and $\mathcal{M}'$ are called diffeomorphic. Diffeomorphic manifolds have the same manifold structure.

1.2.2 Vectors

In pre-relativistic physics as well as in Special Relativity, space and space-time has the structure of a vector space. A vector attached to a point in space-time therefore still 'points' to a new point in the space-time. For a curved manifold this property is lost, the manifold and vector spaces "attached" to it are necessarily separate objects. It is then necessary to consider infinitesimal displacements and to add new mathematical structure to describe how to move between a vector space and the manifold, and how vector spaces at different points are connected.

As for the definition of a manifold, the notion of a tangent vector is intuitive and simple in the case of manifolds embedded in $\mathbb{R}^n$, for example the case of a tangent vector to S^2 embedded in $\mathbb{R}^3$. However, as

[2]More generally one can base the definition of a manifold on the notion of a topological space, but we will neglect this additional structure for the time being.

before we want to give an intrinsic definition of a tangent vector since there is no a priori reason why the true space-time should be embedded in a higher-dimensional space.

In order to do this we consider the link between vectors and directional derivatives. In $\mathbb{R}^n$ it is clear that any vector $\mathbf{v} = (v^1, \ldots, v^n)$ defines a directional derivative operator $\sum_\mu v^\mu \partial/\partial x^\mu$, and vice versa. As derivatives are linear and obey the Leibniz rule for products of functions, we *define* the notion of tangent vector using these requirements:

Let $\mathcal{F}$ denote the ensemble of C^∞ functions $\mathcal{M} \to \mathbb{R}$. We define a tangent vector $\mathbf{v}$ at point $p \in \mathcal{M}$ to be a map $\mathcal{F} \to \mathbb{R}$ which is (1) linear and (2) satisfies the Leibniz rule,

1. $\mathbf{v}(af + bg) = a\mathbf{v}(f) + b\mathbf{v}(g)$ for all $a, b \in \mathbb{R}$ and $f, g \in \mathcal{F}$,
2. $\mathbf{v}(fg) = g[p]\mathbf{v}(f) + f[p]\mathbf{v}(g)$.

Using the addition law $(\mathbf{v}_1 + \mathbf{v}_2)(f) = \mathbf{v}_1(f) + \mathbf{v}_2(f)$ and the product law $(a\mathbf{v})(f) = a\mathbf{v}(f)$, the ensemble of tangent vectors at p, T_p, has the structure of a vector space.

We can explicitly construct a basis of $T_p(\mathcal{M})$ which is a useful exercise. Let $\psi : \mathcal{O} \to U \subset \mathbb{R}^n$ be a chart, with $p \in \mathcal{O}$. If $f \in \mathcal{F}$ then $f \circ \psi^{-1}$ is an application $\mathbb{R}^n \to \mathbb{R}$ that is C^∞ (or *smooth*). For $\mu = 1, \ldots, n$ we can define $\mathbf{X}_\mu : \mathcal{F} \to \mathbb{R}$,

$$\mathbf{X}_\mu(f) = \frac{\partial}{\partial x^\mu} f \circ \psi^{-1}\Big|_{\psi(p)} . \tag{1.18}$$

Here x^μ are the cartesian coordinates of $\mathbb{R}^n$, and the derivative is just the usual partial derivative of a function $\mathbb{R}^n \to \mathbb{R}$. $\mathbf{X}_1$ to $\mathbf{X}_n$ are n linearly independent tangent vectors. That they are linearly independent can be seen e.g. by checking whether $\sum_\mu a^\mu \mathbf{X}_\mu = 0$ implies that all $a^\mu = 0$. If we consider for $f \circ \psi^{-1}$ the functions $\tilde{f}^\nu : (x^1, \ldots, x^n) \to x^\nu$, then we have that

$$a^\nu = \sum_\mu a^\mu \frac{\partial}{\partial x^\mu}(\tilde{f}^\nu) = 0 \tag{1.19}$$

for all ν. This shows that the $\mathbf{X}_\mu$ are linearly independent. Thus the dimension of T_p is at least n. It is also possible to show that the X_μ span T_p, i.e. that they provide an explicit basis, and thus that $\mathrm{dim} T_p(\mathcal{M}) = n = \mathrm{dim}\mathcal{M}$. (You will do this in the exercises.)

The basis vectors $\mathbf{X}_\mu$ are often written $\partial/\partial x^\mu$ or $\boldsymbol{\partial}_\mu$. They depend on the chart ψ (the coordinate system) used, and $\{\mathbf{X}_\mu\}$ is called a *coordinate basis*. For a different chart ψ' we would have found different basis vectors $\{\mathbf{X}'_\nu\}$. Using the chain rule we have

$$\mathbf{X}_\mu = \sum_{\nu=1}^n \frac{\partial x'^\nu}{\partial x^\mu}\Big|_{\psi(p)} \mathbf{X}'_\nu , \tag{1.20}$$

where x'^ν is the ν-th component of the map $\psi' \circ \psi^{-1}$. The components v'^ν of a vector $\mathbf{v}$ are thus related to the components v^μ in the old basis by the vector transformation law

$$v'^\nu = \sum_{\mu=1}^n v^\mu \frac{\partial x'^\nu}{\partial x^\mu} . \tag{1.21}$$

We have now seen that the tangent space forms a vector space of the same dimension as the manifold and that an explicit set of basis vectors $\mathbf{X}_\mu$ can be constructed for each coordinate system. To motivate better why this vector space is called the tangent space we look at the tangent vector of a smooth curve. A smooth curve C on $\mathcal{M}$ is a C^∞ map of (an interval on) $\mathbb{R}$ into $\mathcal{M}$, $C : \mathbb{R} \to \mathcal{M}$, $t \mapsto C(t)$. At each point $p \in \mathcal{M}$ that lies on the curve C we can associate with C a tangent vector $\mathbf{T} \in T_p$: for each $f \in \mathcal{F}$ we set $\mathbf{T}(f)$ equal to the derivative of the function $f \circ C : \mathbb{R} \to \mathbb{R}$, evaluated at p, $\mathbf{T}(f) = d(f \circ C)/dt$. For a given coordinate system ψ the curve $C(t)$ on $\mathcal{M}$ is mapped into a curve $x^\mu(t)$ in $\mathbb{R}^n$ through $\psi \circ C$, and we have that

$$\mathbf{T}(f) = \frac{d}{dt}(f \circ C) = \frac{d}{dt}\left[(f \circ \psi^{-1}) \circ (\psi \circ C)\right] = \sum_\mu \frac{\partial}{\partial x^\mu}\left(f \circ \psi^{-1}\right)\frac{dx^\mu}{dt} = \sum_\mu \frac{dx^\mu}{dt}\mathbf{X}_\mu(f)\,. \quad (1.22)$$

Here we used that $\psi \circ C$ is a function $\mathbb{R} \to \mathbb{R}^n$ and $f \circ \psi^{-1}$ is a function $\mathbb{R}^n \to \mathbb{R}$, so that we can simply use the standard chain rule, and then (1.18). Thus, in a coordinate basis the components T^μ of the tangent vector to a curve $x^\mu(t)$ are given by

$$T^\mu = \frac{dx^\mu}{dt}\,. \quad (1.23)$$

By considering specifically curves $C(t)$ that follow the coordinate directions, like $x^\mu(t) = t\delta^\mu_\nu$ for the direction ν, we see that the tangent vectors to these curves are precisely the basis vectors $\mathbf{X}_\nu$. In other words, these basis vectors 'point' indeed in the coordinate directions.

Tangent fields

There is a tangent space at each point p on $\mathcal{M}$. But there is no general way to relate the tangent spaces $T_p(\mathcal{M})$ and $T_q(\mathcal{M})$ without introducing additional structure, which we will do later. We can however introduce a *tangent field* $\mathbf{v}$ on $\mathcal{M}$, which is the assignment of a tangent vector $\mathbf{v}[p] \in T_p(\mathcal{M})$ at each point $p \in \mathcal{M}$. Although tangent spaces at different points are different, we can nonetheless consider the notion of a smooth tangent field. If f is a C^∞ function on $\mathcal{M}$ then $\mathbf{v}[p](f)$ is a number, i.e. $\mathbf{v}(f)$ is a function on $\mathcal{M}$. The tangent field $\mathbf{v}$ is called smooth if for each smooth function f, the function $\mathbf{v}(f)$ is also smooth. It is easy to verify that the coordinate basis fields $\mathbf{X}_\mu$ are smooth from (1.18), therefore a vector field $\mathbf{v}$ is smooth if and only if its coordinate basis components, v^μ, are smooth functions.

We can associate the tangent fields with transformations of $\mathcal{M}$. A one-parameter group of diffeomorphisms ϕ_t of $\mathcal{M}$ is a C^∞ map $\mathbb{R} \times \mathcal{M} \to \mathcal{M}$ so that for a fixed $t \in \mathbb{R}$, $\phi_t : \mathcal{M} \to \mathcal{M}$ is a diffeomorphism and that for all t and s we have $\phi_t \circ \phi_s = \phi_{t+s}$ (which implies that ϕ_0 is the identity map). For a fixed $p \in \mathcal{M}$, $\phi_t(p) : \mathbb{R} \to \mathcal{M}$ is a curve, the orbit of ϕ_t, which passes through p at $t = 0$. We can now associate a vector field to this one-parameter group of diffeomorphisms by defining $\mathbf{v}[p]$ to be the tangent vector to the orbit of ϕ_t at $t = 0$. This vector field is then the infinitesimal generator of these transformations.

We can also cast the action of a tangent vector on a function, $\mathbf{v}(f)$, in terms of the orbit:

$$\mathbf{v}(f) = \lim_{t \to 0} \frac{f[\phi_t(p)] - f[p]}{t}\,, \quad (1.24)$$

for $\mathbf{v}$ the tangent vector to the orbit of ϕ_t at $t = 0$.

We can also ask the reverse question: given a smooth vector field $\mathbf{v}$ on $\mathcal{M}$, is it possible to find integral curves of $\mathbf{v}$? In a coordinate basis, this is equivalent to finding solutions to the equations

$$\frac{dx^\mu}{dt} = v^\mu(x^1, \ldots, x^n)\,. \quad (1.25)$$

In general such a system of equations has a unique solution, given a starting point at $t = 0$. Therefore to each smooth vector field $\mathbf{v}$ we can associate a unique family of integral curves. For each $p \in \mathcal{M}$ we define $\phi_t(p)$ as the point lying at parameter value t along the integral curve of v starting at p for $t = 0$. It is however possible that a curve extends only to a finite value of t.

Commutator of vector fields

Given two smooth vector fields $\mathbf{u}$ and $\mathbf{v}$ we can define a new vector field $[\mathbf{u},\mathbf{v}]$ called the *commutator* of $\mathbf{u}$ and $\mathbf{v}$, through

$$[\mathbf{u},\mathbf{v}](f) = \mathbf{u}(\mathbf{v}(f)) - \mathbf{v}(\mathbf{u}(f))\,, \tag{1.26}$$

since the commutator satisfies the linearity and Leibniz rule conditions. In $\mathbb{R}^n$ the commutator of two partial derivatives vanishes, so that from (1.18) we have $[\boldsymbol{\partial}_\mu, \boldsymbol{\partial}_\nu] = 0$. The commutator of two vector fields $\mathbf{u}$ and $\mathbf{v}$ thus can be written as

$$[\mathbf{u},\mathbf{v}] = \left[\sum_\mu u^\mu \boldsymbol{\partial}_\mu, \sum_\nu v^\nu \boldsymbol{\partial}_\nu\right] = \sum_{\mu,\nu}(u^\mu \partial_\mu v^\nu - v^\mu \partial_\mu u^\nu)\boldsymbol{\partial}_\nu\,. \tag{1.27}$$

It is also possible to show that for a collection of (non-vanishing) vector fields $\mathbf{X}_1$ to $\mathbf{X}_n$ that commute and are linearly independent at each point one can find a chart for which they are the coordinate basis vector fields.

By construction, the commutator of three smooth vector fields $\mathbf{u},\mathbf{v},\mathbf{w}$ satisfies the Jacobi identity,

$$[[\mathbf{u},\mathbf{v}],\mathbf{w}] + [[\mathbf{v},\mathbf{w}],\mathbf{u}] + [[\mathbf{w},\mathbf{u}],\mathbf{v}] = 0\,. \tag{1.28}$$

1.2.3 Tensors and notational aspects

The notion of a tensor is a direct generalisation of the notion of a vector. Here we will be relatively concise, more details can be found e.g. in the lecture notes for mathematical methods II (MM-II) and in the exercises. Tensors generally play an important role in physics, and especially in GR as we will soon see.

Tensors are “geometric” objects that are independent of the choice of basis in which they are expressed. The simplest example is that of a vector. The components of a vector change if we change basis, but the vector itself does not change. This property leads to a specific relationship between the components of a tensor in different bases, a point that was stressed in MM-II, and that we will only briefly mention below.

Due to their geometric nature, one really should denote a vector or tensor as a coordinate-independent object, e.g. $\mathbf{v}$ for a vector, while v^μ would be the coordinates of $\mathbf{v}$ for a specific choice of basis vectors $\mathbf{e}_\mu$. However, as discussed in MM-II and also below, there are two types of vectors and tensors, covariant and contravariant (and more for tensors). The type is usually distinguished by the placement of the indices. Because of this, we will often write v^μ for the contravariant vector $\mathbf{v}$ even without having chosen a basis, and e.g. $T^{\alpha\beta}_\gamma$ for a tensor that is twice contravariant and one time covariant. In a notational departure from [1] we will normally use Greek characters like μ for space-time indices that run from 0 to 3, and latin characters (e.g. j) for purely spatial indices (running from 1 to 3), except where noted differently. We will also usually use the Einstein summation convention, i.e. repeated indices are summed over. So in general an object denoted $X^{\alpha\beta\dots}_{\gamma\delta\dots}$ will be another way to write the tensor $\mathbf{X}$ that carries additional information; e.g. $X^{\alpha\beta}v_\beta$ would be a vector (which would be difficult to write in abstract notation). The components of $\mathbf{X}$, once a basis has been chosen, will be written in the same way, but usually it is clear from the context whether we mean the abstract tensor or its components. We will occasionally encounter objects that are written in this form without being tensors, like the Christoffel symbols $\Gamma^\alpha_{\mu\nu}$ that we will introduce later, but this will be rare and we will in these cases stress that these are not tensors.

We start with the notion of the dual vector space to a vector space V. The elements in the dual vector space are linear maps $f : V \to \mathbb{R}$. If the set of n vectors $\{\mathbf{e}_\mu\}$ (where now μ enumerates the n vectors) is a basis of V then we can construct a dual basis of V^* through the linear maps $\mathbf{e}^{*\nu}$ (where again ν enumerates the dual basis vectors),

$$\mathbf{e}^{*\nu}(\mathbf{e}_\mu) = \delta^\nu_\mu\,. \tag{1.29}$$

Thanks to the linearity of the elements of V^* this relation fully determines all elements in V^* as they can be written as a linear combination of the dual basis vectors $\mathbf{e}^{*\mu}$. We see that $\dim V^* = \dim V$. The dual-dual vector space V^{**} can be identified with V through a natural isomorphism (see MM-II). Through the relation $\mathbf{e}^{*\mu} \leftrightarrow \mathbf{e}_\mu$ we can define an isomorphism also between V and V^*, but this isomorphism depends in general on the choice of basis vectors $\mathbf{e}_\mu$. We can specify a unique isomorphism only with the help of additional structure, specifically a metric as we will see below.

As a notational side remark, if we are dealing with the tangent space T_p (and co-tangent space T_p^*), then given a coordinate basis $\partial/\partial x^\mu$, one usually denotes the corresponding dual basis as $\boldsymbol{d}x^\mu$. However, this is just a symbol for the linear maps defined through

$$\boldsymbol{d}x^\mu \left(\frac{\partial}{\partial x^\nu} \right) = \boldsymbol{d}x^\mu(\boldsymbol{\partial}_\nu) = \delta^\mu_\nu \,. \tag{1.30}$$

A tensor then is simply a multilinear map from a product of several V and V^* to $\mathbb{R}$,

$$\mathbf{T} : V^* \times \ldots \times V^* \times V \times \ldots \times V \to \mathbb{R} \,. \tag{1.31}$$

If $\mathbf{T}$ acts on the product of j elements of V^* and k elements of V then it is of type (j,k). It associates to j co-vectors and k vectors a real number, and it is linear in each of its arguments. Hence a tensor of type $(1,0)$ is a “normal” vector (also called a contravariant vector), an element of V, while a tensor of type $(0,1)$ is co-vector (or covariant vector), an element of V^*. The ensemble of all tensors of type (j,k) forms a vector space itself, $\mathcal{T}(j,k)$, of dimension n^{j+k}.

Given two tensors $\mathbf{T}_1$ of type (j_1,k_1) and $\mathbf{T}_2$ of type (j_2,k_2) we can define a new tensor of type (j_1+j_2, k_1+k_2) through the outer product, $\mathbf{T}_1 \otimes \mathbf{T}_2$, where $\mathbf{T}_1$ acts on the first j_1 co-vectors and k_1 vectors, and $\mathbf{T}_2$ on the rest, in the obvious way.

This means that we can construct a tensor of type (j,k) from the outer product of j vectors and k co-vectors. Tensors of this kind are called *simple*. For a basis $\mathbf{e}_\mu$ of V and $\mathbf{e}^\nu$ of V^* we have that the n^{j+k} simple tensors $\mathbf{e}_{\mu_1} \otimes \cdots \otimes \mathbf{e}_{\mu_j} \otimes \mathbf{e}^{*\nu_1} \otimes \cdots \otimes \mathbf{e}^{*\nu_k}$ form a basis of $\mathcal{T}(j,k)$. Any tensor $\mathbf{T}$ of type (j,k) can then be expressed in this basis,

$$\mathbf{T} = T^{\mu_1 \ldots \mu_j}_{\nu_1 \ldots \nu_k} \mathbf{e}_{\mu_1} \otimes \cdots \otimes \mathbf{e}_{\mu_j} \otimes \mathbf{e}^{*\nu_1} \otimes \cdots \otimes \mathbf{e}^{*\nu_k} \tag{1.32}$$

where we used Einstein’s summation convention, i.e. all indices are summed over. $T^{\mu_1 \ldots \mu_j}_{\nu_1 \ldots \nu_k}$ are the components of $\mathbf{T}$ in the basis $\{\mathbf{e}_\mu\}$, but as mentioned above, we will usually denote the basis-independent tensor $\mathbf{T}$ by $T^{\mu_1 \ldots \mu_j}_{\nu_1 \ldots \nu_k}$ without specifying a basis. The advantage is that this notation immediately tells us the type of tensor that we are dealing with.

An additional advantage of this notation is that it simplifies an important operation for tensors, called contraction. Contraction is a map $\mathcal{T}(j,k) \to \mathcal{T}(j-1,k-1)$. In coordinates, if we contract upper index l with lower m we get

$$T^{\mu_1 \ldots \mu_l \ldots \mu_j}_{\nu_1 \ldots \nu_m \ldots \nu_k} \to T^{\mu_1 \ldots \mu_{j-1}}_{\nu_1 \ldots \nu_{k-1}} = T^{\mu_1 \ldots \alpha \ldots \mu_j}_{\nu_1 \ldots \alpha \ldots \nu_k} \,. \tag{1.33}$$

In other words, we sum over the contracted indices. Although we wrote the contraction in coordinate form, the resulting quantity is indeed a tensor of lower rank, independent of the basis chosen.

For a tensor on the (co-)tangent space, the vector transformation law (1.21) generalises as follows: Firstly for the dual vectors, if ω_μ are the components of a co-vector with respect to the dual basis $\boldsymbol{d}x^\mu$ then, in order for (1.29) to hold, the components transform under a change to a new basis $\boldsymbol{d}x'^\mu$ as

$$\omega'_\nu = \omega_\mu \frac{\partial x^\mu}{\partial x'^\nu} \tag{1.34}$$

where we again used the summation convention (we will from now on stop to mention this fact). For a tensor of type (j,k) each index transforms with the appropriate factor for a vector or co-vector, so that

$$T'^{\mu'_1\ldots\mu'_j}_{\nu'_1\ldots\nu'_k} = T^{\mu_1\ldots\mu_j}_{\nu_1\ldots\nu_k} \frac{\partial x'^{\mu'_1}}{\partial x^{\mu_1}} \cdots \frac{\partial x'^{\mu'_j}}{\partial x^{\mu_j}} \frac{\partial x^{\nu_1}}{\partial x'^{\nu'_1}} \cdots \frac{\partial x^{\nu_k}}{\partial x'^{\nu'_k}} . \tag{1.35}$$

This is called the tensor transformation law, and it ensures that the tensor itself is independent of the choice of basis.

So far we have only discussed tensors in a single vector space. In general we have however a different tangent space T_p at each point p of the manifold. A tensor over each T_p is then a tensor field in the same way as for the vector field discussed earlier. The notion of smoothness transfers easily based on the definitions of smoothness of functions and contravariant vector fields: A covariant vector field $\boldsymbol{\omega}$ is called smooth (C^∞) if for each smooth vector field v the function $\boldsymbol{\omega}(\mathbf{v})$ is smooth. A tensor field of type (j,k) is called smooth if for any smooth covariant vector fields $\boldsymbol{\omega}^1,\ldots,\boldsymbol{\omega}^j$ and contravariant vector fields $\mathbf{v}_1,\ldots,\mathbf{v}_k$ the functions $\mathbf{T}(\boldsymbol{\omega}^1,\ldots,\boldsymbol{\omega}^j,\mathbf{v}_1,\ldots,\mathbf{v}_k)$ are smooth. Expressed in a coordinate basis, a tensor field is thus smooth if its components are smooth functions (and thanks to the tensor transformation law this property does not depend on the specific choice of coordinates). If two tensor fields are smooth then also their sum (if they are of the same rank) and product are smooth.

Finally, it is useful to define symmetric and anti-symmetric parts of tensor fields. For a tensor of type $(0,2)$ we define

$$T_{(\mu\nu)} = \frac{1}{2}\left(T_{\mu\nu} + T_{\nu\mu}\right) , \tag{1.36}$$

$$T_{[\mu\nu]} = \frac{1}{2}\left(T_{\mu\nu} - T_{\nu\mu}\right) . \tag{1.37}$$

In general we define totally symmetric and anti-symmetric tensor fields of type $(0,k)$ as

$$T_{(\mu_1\ldots\mu_k)} = \frac{1}{k!}\sum_\pi T_{\pi(\mu_1)\ldots\pi(\mu_k)} , \tag{1.38}$$

$$T_{[\mu_1\ldots\mu_k]} = \frac{1}{k!}\sum_\pi \sigma_\pi T_{\pi(\mu_1)\ldots\pi(\mu_k)} , \tag{1.39}$$

where the sum runs over all permutations π of k numbers, and where σ_π is $+1$ for an even permutation and -1 for an odd permutation. It is also possible to apply these definitions to contravariant (upper) indices and to sub-sets of indices.

A totally antisymmetric 'tangent' tensor field of type $(0,k)$, i.e. one for which $T_{\mu_1\ldots\mu_k} = T_{[\mu_1\ldots\mu_k]}$, is called a differential k-form – covariant tangent vector fields that are "tensors" of type $(0,1)$ are all differential one-forms. We will defer more detailed discussions of differential forms until we need them.

1.2.4 The metric

The metric or metric tensor is special tensor of fundamental importance in GR. Intuitively a metric should provide a notion of (infinitesimal, squared) distance. We have seen that tangent vector fields, as infinitesimal generators of a one-parameter group of diffeomorphisms, provide a notion of "infinitesimal displacement". It is therefore natural that the metric is a tensor field $\boldsymbol{g}$ of type $(0,2)$ that provides a linear map of $T_p \times T_p$ into real numbers. In addition, a metric needs to be symmetric, i.e. $\boldsymbol{g}(\mathbf{v}_1,\mathbf{v}_2) = \boldsymbol{g}(\mathbf{v}_2,\mathbf{v}_1)$, and non-degenerate (if $\boldsymbol{g}(\mathbf{v},\mathbf{v}_1) = 0$ for all $\mathbf{v} \in T_p$ then $\mathbf{v}_1 = 0$).

In a coordinate basis we can write the metric tensor as

$$\boldsymbol{g} = g_{\mu\nu}\boldsymbol{d}x^\mu \otimes \boldsymbol{d}x^\nu . \tag{1.40}$$

Often one writes ds^2 instead of $\boldsymbol{g}$, and omits the outer product sign, so that an equivalent way to write the above is

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu \,. \tag{1.41}$$

Given a metric, we can always find a (not unique) orthonormal basis $\{\mathbf{e}_\mu\}$, i.e. a basis so that $\boldsymbol{g}(\mathbf{e}_\mu, \mathbf{e}_\nu) = 0$ if $\mu \neq \nu$ and $\boldsymbol{g}(\mathbf{e}_\mu, \mathbf{e}_\mu) = \pm 1$. The number of $+1$ and -1 is independent of the choice of orthonormal basis and is called the *signature* of the metric. A metric with signature $+, \ldots, +$ (i.e. one that is positive definite) is called Riemannian, while one with signature that has one $-$ and the remainder $+$ as in special and general relativity is called Lorentzian.

The metric provides a unique (basis independent) map between T_p and T_p^*, by associating to each contravariant vector $\mathbf{v}$ the covariant vector $\mathbf{v}^* = \boldsymbol{g}(\cdot, \mathbf{v})$. Since $\boldsymbol{g}$ is non-degenerate, this map is one-to-one and onto, an in principle allows to identify T_p and T_p^*. Equivalently we can say that the metric allows to "raise and lower" indices,

$$v_\mu^* = g_{\mu\nu} v^\nu \,. \tag{1.42}$$

This can be done not only for vectors but generally for tensors. For the metric itself we have that

$$g^{\mu\alpha} g_{\alpha\nu} = \delta^\mu_\nu \,. \tag{1.43}$$

1.2.5 Covariant derivative and parallel transport

As mentioned, the tangent spaces at different points are different, there is no canonical way to relate vectors in different tangent spaces, and to e.g. answer the question whether two vectors in T_p and T_q are the same. If we had an additional structure that allows to transport vectors in a parallel way from one tangent space to another, then we can perform such a comparison. We can then also define a derivative of vector fields in the direction of a curve, by parallel transporting one vector along the curve and looking at the difference to the vector of the vector field at that point. Conversely, the ability to compute the derivative of a vector field would confer the ability to define parallel transport by defining that a vector is transported in a parallel way if its derivative is zero. The two concepts are thus closely related. In addition, as we will discuss in the next subsection, by looking at parallel transport along closed curves we can also define a notion of curvature that is intrinsic to a manifold and does not require any embedding in a higher-dimensional space.

We will start by constructing a convenient derivative operator. Convenient means that it should behave like a derivative, i.e. acting linearly on its argument and obeying the Leibniz rule, that it should reduce to the usual partial derivative in the limit of flat space with a coordinate basis, but that the derivative of a tensor should transform like a tensor on an arbitrary manifold. A derivative ∇_μ would associate to a vector field X^ν an object $\nabla_\mu X^\nu$ with two indices, i.e. a tensor $(1,1)$,

$$(\nabla \mathbf{X})(\mathbf{Y}, \omega) = \omega(\nabla_{\mathbf{Y}} \mathbf{X}) \in \mathbb{R} \,. \tag{1.44}$$

The action of $\nabla \mathbf{X}$ on another vector field $\mathbf{Y}$ is thus a third vector field $\mathbf{Z} = \nabla_{\mathbf{Y}} \mathbf{X}$, in other words a derivative operator can be considered both as a map that associates to two vector fields $\mathbf{X}$ and $\mathbf{Y}$ a third vector field $\mathbf{Z}$, or as a map from a tensor field $(1,0)$ (a vector field) to a tensor field $(1,1)$. More generally, a derivative operator ∇ should be a map from smooth (or at least differentiable) tensor fields (k,l) to smooth tensor fields $(k, l+1)$, satisfying (for tensors $\mathbf{T}$ and $\mathbf{S}$)

$$\nabla(\mathbf{T} + \mathbf{S}) = \nabla \mathbf{T} + \nabla \mathbf{S} \,, \tag{1.45}$$

$$\nabla(\mathbf{T} \otimes \mathbf{S}) = (\nabla \mathbf{T}) \otimes \mathbf{S} + \mathbf{T} \otimes (\nabla \mathbf{S}) \,. \tag{1.46}$$

(Note that for the linearity property $\mathbf{T}$ and $\mathbf{S}$ must be tensors of the same type (k,l) while for the Leibniz rule they can in general be of different type.) As the derivative of functions defined through (1.24) is fairly canonical, we want that the derivative operator along a tangent vector field $\mathbf{v}$, applied to a function, to be simply

$$\nabla_{\mathbf{v}}(f) = \mathbf{v}(f)\,, \tag{1.47}$$

to ensure that we can continue to consider tangent vectors as directional derivatives on scalar fields. We also want the derivative operator to be 'f-linear' in its first argument,

$$\nabla_{\mathbf{u}+f\mathbf{v}}\mathbf{w} = \nabla_{\mathbf{u}}\mathbf{w} + f\nabla_{\mathbf{v}}\mathbf{w} \tag{1.48}$$

for vector fields $\mathbf{u}, \mathbf{v}, \mathbf{w}$ and $f \in \mathcal{F}$. The f-linearity is important to ensure that we can consistently write

$$\mathbf{v}(f) = \nabla_{\mathbf{v}}(f) = v^\mu \nabla_{\boldsymbol{\partial}_\mu} f = v^\mu \nabla_\mu f = v^\mu \boldsymbol{\partial}_\mu(f)\,. \tag{1.49}$$

We also introduced the notation ∇_μ as shorthand for $\nabla_{\boldsymbol{\partial}_\mu}$, which is what was implicitly meant above.

Let us consider the action of a derivative operator on a coordinate basis $\boldsymbol{\partial}_\mu$. As it associates to two vector fields a third vector field, we can write in a coordinate basis

$$\nabla_{\boldsymbol{\partial}_\mu}\boldsymbol{\partial}_\nu = \Gamma^\rho_{\mu\nu}\boldsymbol{\partial}_\rho\,. \tag{1.50}$$

Here the set of numbers $\Gamma^\rho_{\mu\nu}$ are called *connection coefficients* or *Christoffel symbols*. The object Γ is not a tensor, you will compute in the exercises how it has to transform under coordinate changes, in order to ensure that $\nabla\mathbf{T}$ *is* a tensor. You should find

$$\Gamma'^\alpha_{\beta\gamma} = \frac{\partial x^\mu}{\partial x'^\beta}\frac{\partial x^\nu}{\partial x'^\gamma}\frac{\partial x'^\alpha}{\partial x^\rho}\Gamma^\rho_{\mu\nu} + \frac{\partial^2 x^\rho}{\partial x'^\beta \partial x'^\gamma}\frac{\partial x'^\alpha}{\partial x^\rho}\,. \tag{1.51}$$

Applying (1.50) to a vector field $\mathbf{v} = v^\mu \boldsymbol{\partial}_\mu$ we find

$$\nabla_\mu \mathbf{v} = \nabla_{\boldsymbol{\partial}_\mu}(v^\nu \boldsymbol{\partial}_\nu) = (\partial_\mu v^\nu)\boldsymbol{\partial}_\nu + \Gamma^\rho_{\mu\nu}v^\nu\boldsymbol{\partial}_\rho = \left(\partial_\mu v^\nu + \Gamma^\nu_{\mu\rho}v^\rho\right)\boldsymbol{\partial}_\nu\,. \tag{1.52}$$

We see that if the Christoffel symbols vanish, the derivative operator becomes just a partial derivative.

In addition to the properties mentioned above we would like a derivative operator to commute with contractions, so that (here illustrated for a $(1,2)$ tensor $\mathbf{T}$, but of course we want to hold this for general (k,l) tensors),

$$\nabla_\mu\left(T^\nu_{\nu\rho}\right) = \left(\nabla_\mu T\right)^\nu_{\nu\rho}\,. \tag{1.53}$$

Why is this useful? A priori a covariant vector field $\boldsymbol{\omega}$, while verifying an analogous equation to (1.52), does not need to lead to the same Christoffel symbols,

$$\nabla_\mu\boldsymbol{\omega} = \nabla_\mu \omega_\nu \boldsymbol{d}x^\nu = (\partial_\mu\omega_\nu)\boldsymbol{d}x^\nu + \tilde{\Gamma}^\nu_{\mu\rho}\omega_\nu\boldsymbol{d}x^\rho\,. \tag{1.54}$$

Let us apply the covariant derivative to the function $\boldsymbol{\omega}(\mathbf{v}) = \omega_\mu v^\mu$ which we can now consider as the contraction of $\boldsymbol{\omega}\otimes\mathbf{v}$:

$$\begin{aligned}
\nabla_\mu\left(\omega_\lambda v^\lambda\right) &= \left(\nabla_\mu\omega_\lambda\right)v^\lambda + \omega_\lambda\left(\nabla_\mu v^\lambda\right) \\
&= \left(\partial_\mu\omega_\lambda\right)v^\lambda + \tilde{\Gamma}^\sigma_{\mu\lambda}\omega_\sigma v^\lambda + \omega_\lambda\left(\partial_\mu v^\lambda\right) + \omega_\lambda\Gamma^\lambda_{\mu\rho}v^\rho \\
&= \partial_\mu\left(\omega_\lambda v^\lambda\right) = \left(\partial_\mu\omega_\lambda\right)v^\lambda + \omega_\lambda\left(\partial_\mu v^\lambda\right)\,,
\end{aligned} \tag{1.55}$$

where in the last equality we used the property (1.47) for $\mathbf{v} = \boldsymbol{\partial}_\mu$. As $\boldsymbol{\omega}$ and $\mathbf{v}$ were arbitrary we conclude that

$$\tilde{\Gamma}^\sigma_{\mu\lambda} = -\Gamma^\sigma_{\mu\lambda}\,. \tag{1.56}$$

The covariant derivative of a covariant vector field is thus based on the same Christoffel symbols as the covariant derivative of a contravariant vector field, up to the sign,

$$\nabla_\mu \omega_\nu = (\partial_\mu \omega_\nu) - \Gamma^\lambda_{\mu\nu} \omega_\lambda \,. \tag{1.57}$$

For a general tensor $\mathbf{T}$ of type (k, l) we then have that

$$\begin{aligned} \nabla_\lambda T^{\mu_1 \dots \mu_k}_{\nu_1 \dots \nu_l} &= \partial_\lambda T^{\mu_1 \dots \mu_k}_{\nu_1 \dots \nu_l} \\ &\quad + \Gamma^{\mu_1}_{\lambda\rho} T^{\rho\mu_2 \dots \mu_k}_{\nu_1 \dots \nu_k} + \Gamma^{\mu_2}_{\lambda\rho} T^{\mu_1 \rho \dots \mu_k}_{\nu_1 \dots \nu_k} + \cdots \\ &\quad - \Gamma^{\rho}_{\lambda\nu_1} T^{\mu_1 \dots \mu_k}_{\rho\nu_2 \dots \nu_l} - \Gamma^{\rho}_{\lambda\nu_2} T^{\mu_1 \dots \mu_k}_{\nu_1 \rho \dots \nu_l} - \cdots \,. \end{aligned} \tag{1.58}$$

For a derivative operator ∇_μ we can define *parallel transport* of a vector field v^ν along a curve $C(t)$ (given by $x^\mu(t)$ in coordinates) with tangent $t^\alpha = dx^\mu(t)/dt$ through the requirement that

$$\nabla_{\mathbf{t}} \mathbf{v} = t^\mu \nabla_\mu v^\nu = 0 \tag{1.59}$$

is satisfied along the curve. More generally, a tensor field is parallel transported if it satisfies the same equation. In a coordinate system, the equation for parallel transport of a vector becomes

$$t^\mu \partial_\mu v^\nu + t^\mu \Gamma^\nu_{\mu\alpha} v^\alpha = \frac{dv^\nu}{dt} + t^\mu \Gamma^\nu_{\mu\alpha} v^\alpha = 0 \,. \tag{1.60}$$

We can see that the parallel transport of a vector or tensor only depends on its values along the curve, so that we do not need a full vector or tensor field in order to describe parallel transport. From the differential equation we can also see that a vector at a point p on the curve uniquely defines a parallel transported vector everywhere else on the curve. We can use this to identify (map onto each other) tangent spaces T_p and T_q with a given derivative operator and a curve. Such an identification is also called a connection – often in differential geometry one starts with the definition of a connection and then constructs a derivative operator from it.

In general many derivative operators are possible: we obtain a different one for every choice of connection coefficients. An additional condition that we want to impose here (and generally for GR, although this condition can be dropped for extended theories of gravity) is that the connection should be torsion free,

$$\nabla_\mu \nabla_\nu f = \nabla_\nu \nabla_\mu f \quad \forall f \in \mathcal{F} \,. \tag{1.61}$$

In terms of Christoffel symbols this condition requires symmetry in the lower indices,

$$\Gamma^\rho_{\mu\nu} = \Gamma^\rho_{\nu\mu} \,. \tag{1.62}$$

We note that for a torsion-free connection we can write the commutator of two vector fields (1.27), using (1.49), also as

$$\begin{aligned} [\mathbf{v}, \mathbf{w}](f) &= \mathbf{v}(\mathbf{w}(f)) - \mathbf{w}(\mathbf{v}(f)) = v^\mu \nabla_\mu (w^\nu \nabla_\nu f) - w^\mu \nabla_\mu (v^\nu \nabla_\nu f) \\ &= (v^\mu \nabla_\mu w^\nu - w^\mu \nabla_\mu v^\nu) \nabla_\nu f \end{aligned} \tag{1.63}$$

which implies that

$$[\mathbf{v}, \mathbf{w}]^\nu = (v^\mu \nabla_\mu w^\nu - w^\mu \nabla_\mu v^\nu) \,. \tag{1.64}$$

Given a metric, there is a unique choice of derivative operator that preserves the inner product defined by the metric[3]. What we mean with this is that, for two parallel-transported vectors v^μ and w^ν,

$$t^\alpha \nabla_\alpha (g_{\mu\nu} v^\mu w^\nu) = 0 \,. \tag{1.65}$$

[3] However, in general a metric is not necessary to define a connection! This is obvious since we have not used the metric until now.

Since v and w are parallel transported, using the Leibniz rule we find that this requires the additional condition that $t^\alpha v^\mu w^\nu \nabla_\alpha g_{\mu\nu} = 0$. This will hold for all curves and parallel-transported vectors if and only if

$$\nabla_\alpha g_{\mu\nu} = 0\,. \tag{1.66}$$

This condition determines the derivative operator uniquely. We can show this by explicitly constructing it in a coordinate basis, which can be done by considering the equation for metric compatibility for different permutations of the indices,

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

We subtract last two from the first one and use the symmetry of the Christoffel symbols and the metric to find

$$\partial_\rho g_{\mu\nu} - \partial_\mu g_{\nu\rho} - \partial_\nu g_{\rho\mu} + 2\Gamma^\lambda_{\mu\nu} g_{\lambda\rho} = 0\,. \tag{1.67}$$

By multiplying with the inverse metric we can solve for the Christoffel symbols,

$$\Gamma^\rho_{\mu\nu} = \frac{1}{2} g^{\rho\sigma}\left(\frac{\partial g_{\nu\sigma}}{\partial x^\mu} + \frac{\partial g_{\mu\sigma}}{\partial x^\nu} - \frac{\partial g_{\mu\nu}}{\partial x^\sigma}\right)\,. \tag{1.68}$$

From now on we will always use this connection (also known as Levi-Civita connection) in these lectures. A notational point: We will occasionally write, for an arbitrary quantity A, $A_{,\mu}$ for the partial derivative $\partial_\mu A$, and $A_{;\mu}$ for $\nabla_\mu A$.

1.2.6 Curvature

For a surface embedded in three-dimensional space, like a two-sphere, it appears immediately obvious whether it is curved or not.[4] However, the notion of curvature is much less simple when we try to define it in an intrinsic way, without an embedding in a higher-dimensional space. A key ingredient is parallel transport:

If we parallel transport a vector in a flat two dimensional plane around a closed curve (e.g. a triangle) then the vector will remain parallel after the round trip. But consider the case of a two-sphere, Fig. 1.1. We parallel transport a vector from point A pointing towards the north pole to N, then to point B and finally along the equator back to point A. We find that the vector has rotated by an angle that depends on the surface of the triangle described by the path along which we parallel transported the vector.

To consider parallel transport along an infinitesimal loop, we study the commutator of two derivative operations, $[\nabla_\mu, \nabla_\nu]$, based on the argument that this should correspond to an infinitesimal parallelogram in the directions $\boldsymbol{\partial}_\mu$ and $\boldsymbol{\partial}_\nu$. It is maybe simplest to just calculate the explicit expression, for a vector field v^μ, using Eq. (1.52):

$$\begin{aligned}
[\nabla_\mu, \nabla_\nu] v^\rho &= \nabla_\mu \nabla_\nu v^\rho - \nabla_\nu \nabla_\mu v^\rho\\
&= \partial_\mu \left(\nabla_\nu v^\rho\right) - \Gamma^\lambda_{\mu\nu} \nabla_\lambda v^\rho + \Gamma^\rho_{\mu\sigma} \nabla_\nu v^\sigma - (\mu \leftrightarrow \nu)\\
&= \partial_\mu \partial_\nu v^\rho + (\partial_\mu \Gamma^\rho_{\nu\sigma}) v^\sigma + \Gamma^\rho_{\nu\sigma} \partial_\mu v^\sigma - \Gamma^\lambda_{\mu\nu} \partial_\lambda v^\rho - \Gamma^\lambda_{\mu\nu} \Gamma^\rho_{\lambda\sigma} v^\sigma + \Gamma^\rho_{\mu\sigma} \partial_\nu v^\sigma + \Gamma^\rho_{\mu\sigma} \Gamma^\sigma_{\nu\lambda} v^\lambda\\
&\quad -(\mu \leftrightarrow \nu)\\
&= \left(\partial_\mu \Gamma^\rho_{\nu\sigma} - \partial_\nu \Gamma^\rho_{\mu\sigma} + \Gamma^\rho_{\mu\lambda} \Gamma^\lambda_{\nu\sigma} - \Gamma^\rho_{\nu\lambda} \Gamma^\lambda_{\mu\sigma}\right) v^\sigma\,.
\end{aligned} \tag{1.69}$$

[4] Care needs to be taken to avoid being fooled by the difference between intrinsic and extrinsic curvature. For example, a torus is actually flat intrinsically even though it may look curved when embedded in a higher-dimensional space.

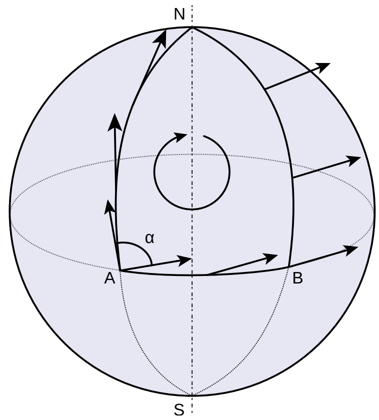

Figure 1.1: Parallel transport of a vector around a closed loop, from A to N to B and back to A, on the sphere. Image from Wikipedia (Fred the Oyster, CC BY-SA 4.0 license, unmodified).

In the last step we have used the symmetry of the Christoffel symbols for a torsion free connection, otherwise an additional term (involving another tensor, the torsion tensor) would appear. As the left hand side is manifestly a tensor, we see that the object in brackets must be tensor too, and we can write

$$[\nabla_\mu, \nabla_\nu] v^\rho = R^\rho_{\sigma\mu\nu} v^\sigma \,. \tag{1.70}$$

The new tensor $\mathbf{R}$ is called the *Riemann curvature tensor*, and from the above calculation we can see that its components are given by

$$R^\rho_{\sigma\mu\nu} = \partial_\mu \Gamma^\rho_{\nu\sigma} - \partial_\nu \Gamma^\rho_{\mu\sigma} + \Gamma^\rho_{\mu\lambda}\Gamma^\lambda_{\nu\sigma} - \Gamma^\rho_{\nu\lambda}\Gamma^\lambda_{\mu\sigma} \,. \tag{1.71}$$

It is straightforward to check that its components indeed obey the appropriate tensor transformation law under coordinate changes. We see that the Riemann curvature tensor is only constructed from the connection, the metric does not appear directly. From its construction as the commutator of covariant derivatives, it was also not obvious that its action on a vector field is purely a multiplicative linear transformation. As a tensor, $\mathbf{R}$ should be a multilinear map from three vector fields $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ to a new vector field (and hence of rank $(1,3)$). Specifically, we can write

$$\mathbf{R}(\mathbf{X},\mathbf{Y})\mathbf{Z} = \nabla_\mathbf{X}\nabla_\mathbf{Y}\mathbf{Z} - \nabla_\mathbf{Y}\nabla_\mathbf{X}\mathbf{Z} - \nabla_{[\mathbf{X},\mathbf{Y}]}\mathbf{Z} \,. \tag{1.72}$$

If $\mathbf{X}$ and $\mathbf{Y}$ are coordinate basis fields then their commutator vanishes and we obtain the previous formula. With $[\mathbf{X},\mathbf{Y}] = (X^\mu \nabla_\mu Y^\nu - Y^\mu \nabla_\mu X^\nu)\boldsymbol{\partial}_\nu$ from (1.64) we write (1.72) in components as

$$R^\rho_{\sigma\mu\nu} X^\mu Y^\nu Z^\sigma = X^\mu \nabla_\mu \left(Y^\nu \nabla_\nu Z^\rho\right) - Y^\mu \nabla_\mu \left(X^\nu \nabla_\nu Z^\rho\right) - (X^\mu \nabla_\mu Y^\nu - Y^\mu \nabla_\mu X^\nu)\nabla_\nu Z^\rho \,, \tag{1.73}$$

which we can check to be equivalent to (1.70); we see that $\nabla_{[\mathbf{X},\mathbf{Y}]}$ serves to remove the terms arising from the derivatives acting on the functions X^μ and Y^ν. An explicit calculation also shows that $\mathbf{R}(f\mathbf{X}, g\mathbf{Y})h\mathbf{Z} = fgh\mathbf{R}(\mathbf{X},\mathbf{Y})\mathbf{Z}$ for $f, g, h \in \mathcal{F}(\mathcal{M})$.

To find the action of $[\nabla_\mu, \nabla_\nu]$ on a one-form ω_μ we could explicitly compute the answer as we did above for the vector field v^μ. Alternatively we can apply the commutator of the covariant derivatives to the function $v^\mu \omega_\mu$ since on a function it vanishes for a torsion-free connection that verifies (1.61),

$$\begin{aligned} 0 &= (\nabla_\mu\nabla_\nu - \nabla_\nu\nabla_\mu)(v^\rho\omega_\rho) = \nabla_\mu(\omega_\rho\nabla_\nu v^\rho + v^\rho\nabla_\nu\omega_\rho) - \nabla_\nu(\omega_\rho\nabla_\mu v^\rho + v^\rho\nabla_\mu\omega_\rho) \\ &= \omega_\rho(\nabla_\mu\nabla_\nu - \nabla_\nu\nabla_\mu)v^\rho + v^\rho(\nabla_\mu\nabla_\nu - \nabla_\nu\nabla_\mu)\omega_\rho \\ &= \omega_\rho v^\sigma R^\rho_{\sigma\mu\nu} + v^\rho(\nabla_\mu\nabla_\nu - \nabla_\nu\nabla_\mu)\omega_\rho \,. \end{aligned} \tag{1.74}$$

As $\mathbf{v}$ was arbitrary, we thus find

$$[\nabla_\mu, \nabla_\nu]\omega_\rho = -R^\sigma_{\rho\mu\nu}\omega_\sigma\,. \tag{1.75}$$

We can by induction extend this result to a general tensor field,

$$[\nabla_\mu, \nabla_\nu]T^{\sigma_1\ldots\sigma_j}_{\rho_1\ldots\rho_k} = R^{\sigma_1}_{\lambda\mu\nu}T^{\lambda\sigma_2\ldots\sigma_j}_{\rho_1\ldots\rho_k} + \cdots - R^\lambda_{\rho_1\mu\nu}T^{\sigma_1\ldots\sigma_j}_{\lambda\rho_2\ldots\rho_k} - \cdots\,, \tag{1.76}$$

where all the indices of $\mathbf{T}$ are affected in turn (symbolized by the $\cdots$).

The Riemann tensor has some important properties:

1. $R^\rho_{\sigma\mu\nu}$ is antisymmetric in the last two indices (notice that this depends on the index placement convention), equivalent to $\mathbf{R}(\mathbf{X},\mathbf{Y}) = -\mathbf{R}(\mathbf{Y},\mathbf{X})$. This is obvious from the expression (1.71).

2. It is also antisymmetric in the first two indices, $R_{\rho\sigma\mu\nu} = -R_{\sigma\rho\mu\nu}$ (for the Levi-Civita connection). This can be shown by applying $[\nabla_\mu, \nabla_\nu]$ to the metric,

$$0 = [\nabla_\mu, \nabla_\nu]g_{\rho\sigma} = -R^\lambda_{\rho\mu\nu}g_{\lambda\sigma} - R^\lambda_{\sigma\mu\nu}g_{\rho\lambda} = -\left(R_{\sigma\rho\mu\nu} + R_{\rho\sigma\mu\nu}\right)\,. \tag{1.77}$$

3. $R^\rho_{[\sigma\mu\nu]} = 0$ (which also implies that the sum of the cyclic permutations in σ, μ, ν of $R^\rho_{\sigma\mu\nu}$ vanishes). With a direct calculation using (1.57), the commutativity of the usual partial derivative and the symmetry of the Christoffel symbols (1.62) we can show that $\nabla_{[\mu}\nabla_\nu\omega_{\sigma]} = 0$. Using this result we have

$$0 = 2\nabla_{[\mu}\nabla_\nu\omega_{\sigma]} = \nabla_{[\mu}\nabla_\nu\omega_{\sigma]} - \nabla_{[\nu}\nabla_\mu\omega_{\sigma]} = R^\rho_{[\sigma\mu\nu]}\omega_\rho\,, \tag{1.78}$$

which gives the desired result as $\boldsymbol{\omega}$ was arbitrary. One can also verify the property using the explicit expression (1.71), again using the symmetry of the Christoffel symbols.

4. $R_{\rho\sigma\mu\nu} = R_{\mu\nu\rho\sigma}$ from properties 1, 2 and 3. (Needs a bit of playing around.)

5. The Bianchi identity: $\nabla_{[\lambda}R_{|\rho\sigma|\mu\nu]} = 0$ where the vertical bars indicate that we antisymmetrise over μ, ν and λ but not ρ and σ (although thanks to property 4 we could have written $\nabla_{[\lambda}R_{\rho\sigma]\mu\nu} = 0$, and using the antisymmetry properties of the Riemann tensor we could also have used a cyclic sum instead of the antisymmetrisation). This implies

$$[[\nabla_\lambda, \nabla_\rho], \nabla_\sigma] + [[\nabla_\rho, \nabla_\sigma], \nabla_\lambda] + [[\nabla_\sigma, \nabla_\lambda], \nabla_\rho] = 0\,, \tag{1.79}$$

very similar to the Jacobi identity (1.28). To prove the Bianchi identity we consider

$$-\left(\nabla_\mu\nabla_\nu - \nabla_\nu\nabla_\mu\right)\nabla_\rho\omega_\sigma = R^\lambda_{\rho\mu\nu}\nabla_\lambda\omega_\sigma + R^\lambda_{\sigma\mu\nu}\nabla_\rho\omega_\lambda \tag{1.80}$$

as well as

$$-\nabla_\mu(\nabla_\nu\nabla_\rho\omega_\sigma - \nabla_\rho\nabla_\nu\omega_\sigma) = \nabla_\mu\left(R^\lambda_{\sigma\nu\rho}\omega_\lambda\right) = \omega_\lambda\nabla_\mu R^\lambda_{\sigma\nu\rho} + R^\lambda_{\sigma\nu\rho}\nabla_\mu\omega_\lambda\,. \tag{1.81}$$

If we antisymmetrise these two equations over μ, ν and ρ the left-hand side of both become equal and we thus have

$$R^\lambda_{[\rho\mu\nu]}\nabla_\lambda\omega_\sigma + R^\lambda_{\sigma[\mu\nu}\nabla_{\rho]}\omega_\lambda = \omega_\lambda\nabla_{[\rho}R^\lambda_{|\sigma|\mu\nu]} + R^\lambda_{\sigma[\mu\nu}\nabla_{\rho]}\omega_\lambda \tag{1.82}$$

where the vertical bars around σ indicate that it is not part of the antisymmetrisation. The first term on the left hand side vanishes thanks to property 3, while the second terms on both sides cancel, so that

$$\omega_\lambda\nabla_{[\rho}R^\lambda_{|\sigma|\mu\nu]} = 0 \tag{1.83}$$

for all $\boldsymbol{\omega}$, which proves property 5.

We can decompose the Riemann tensor into a "trace" and a "trace-free" part. Thanks to the anti-symmetry properties of the Riemann tensor with respect to the first two and last two indices, contractions of those indices vanish. The trace over the second and fourth (or first and third) index defines the *Ricci tensor*,

$$R_{\alpha\beta} = R^{\mu}_{\alpha\mu\beta}\,, \tag{1.84}$$

and the trace of the Ricci tensor is the *Ricci scalar*, $R = R^{\alpha}_{\alpha}$ also called *scalar curvature*. The Ricci tensor inherits the symmetry property $R_{\alpha\beta} = R_{\beta\alpha}$ from property 4 above. The trace-free part is called the *Weyl tensor*, it is defined through (for $n \geq 3$ dimensions)

$$R_{\mu\nu\sigma\rho} = C_{\mu\nu\sigma\rho} + \frac{2}{n-2}\left(g_{\mu[\sigma}R_{\rho]\nu} - g_{\nu[\sigma}R_{\rho]\mu}\right) - \frac{2}{(n-1)(n-2)}Rg_{\mu[\sigma}g_{\rho]\nu}\,. \tag{1.85}$$

The Weyl tensor satisfies properties 1, 2 and 3 of the Riemann tensor, and it is trace free on all indices.[5]

Contracting the Bianchi identity we find that

$$\nabla_\alpha R^{\alpha}_{\beta\mu\nu} + \nabla_\nu R_{\beta\mu} - \nabla_\mu R_{\beta\nu} = 0\,. \tag{1.86}$$

Raising ν and contracting β and ν we find

$$\nabla_\alpha R^{\alpha}_{\mu} + \nabla_\beta R^{\beta}_{\mu} - \nabla_\mu R = 0 \tag{1.87}$$

which we can rewrite as

$$\nabla^\alpha G_{\alpha\beta} = 0\,, \quad G_{\alpha\beta} = R_{\alpha\beta} - \frac{1}{2}g_{\alpha\beta}R\,. \tag{1.88}$$

$G_{\alpha\beta}$ is called the Einstein tensor. The fact that it satisfies the contracted Bianchi identity will be important in the construction of a consistent theory of gravity.

1.2.7 Geodesics and the geodesic deviation equation

Geodesics are curves that are "as straight as possible". We define a geodesic as a curve whose tangent vector is parallel transported along itself. The tangent vector T^α then satisfies the equation

$$\nabla_{\mathbf{T}}\mathbf{T} = T^\alpha \nabla_\alpha T^\beta = 0\,. \tag{1.89}$$

In principle our prescription above could lead to $T^\alpha \nabla_\alpha T^\beta = aT^\beta$, but one can always find a reparametrisation of the curve so that (1.89) is verified. This is called an affine parameterisation. In a coordinate basis, the above equation is

$$\frac{dT^\alpha}{dt} + \Gamma^{\alpha}_{\beta\gamma}T^\beta T^\gamma = 0\,, \tag{1.90}$$

and using also expression (1.23) for T^μ we can write

$$\frac{d^2x^\alpha}{dt^2} + \Gamma^{\alpha}_{\beta\gamma}\frac{dx^\beta}{dt}\frac{dx^\gamma}{dt} = 0\,. \tag{1.91}$$

This is a system of n coupled second-order ordinary differential equations. For a given initial value of x^μ and dx^μ/dt a solution always exists. This means that there always exists a unique geodesic through a point $p \in \mathcal{M}$ with a tangent T^α.

[5] Similarly as for deformations of elastic bodies, we can roughly consider the Ricci tensor as describing volume changes due to the deformation of space-time, and the Weyl tensor as describing volume-preserving distortions due to tidal forces or space-time shear. We also note here that the Weyl tensor vanishes identically for manifolds with less than four dimensions – this implies that a GR-like theory of gravity in 2 or 3 space-time dimensions would not work very well. Another interesting point is that for a manifold with four or more dimensions and a vanishing Weyl tensor (and for any two-dimensional manifold) one can map a neighbourhood of any point to flat space with a conformal transformation of the metric, $g \to e^{2f}g$ where f is a smooth function, which is called *(locally) conformally flat*. The Weyl tensor is invariant under such conformal transformations of the metric.

Geodesics as extremal curves

The length of a smooth curve $C(t)$ is given by

$$l = \int \sqrt{g_{\mu\nu} T^\mu T^\nu} dt \,. \tag{1.92}$$

The norm is preserved along a geodesic since $\nabla_{\mathbf{T}}(g_{\mu\nu} T^\mu T^\nu) = 0$. This length is also independent of the parametrisation: if we change to a new parameter $s(t)$ then the tangent vector will be $S^\alpha = (dt/ds) T^\alpha$ and thus the length will be the same.

For a metric with Lorentzian signature, a curve is called space-like if $g_{\mu\nu} T^\mu T^\nu > 0$ everywhere, time-like if $g_{\mu\nu} T^\mu T^\nu < 0$ and null if $g_{\mu\nu} T^\mu T^\nu = 0$. For time-like curves we change the sign in the square root of Eq. (1.92), and we call the resulting length *proper time*,

$$\tau = \int \sqrt{-g_{\mu\nu} T^\mu T^\nu} dt \,. \tag{1.93}$$

The length of null curves is zero. Since for a geodesic the tangent vector is parallel transported and has constant norm, geodesics in a Lorentz manifold cannot change from space-like to time-like or null.

The geodesic is actually the curve that extremises the length between its endpoints, as we will now show. Specifically we will consider a space-like curve and work in a chart. In a coordinate basis we can write (1.92) as

$$l = \int_a^b \sqrt{g_{\mu\nu} \frac{dx^\mu}{dt} \frac{dx^\nu}{dt}} dt \,, \tag{1.94}$$

where $C(a) = p$ and $C(b) = q$ are the end-points of the curve. Finding the equation for the extremal curve is similar to the variational problem in Lagrangian mechanics. The variation of l is

$$\delta l = \int_a^b \left[g_{\mu\nu} \frac{dx^\mu}{dt} \frac{dx^\nu}{dt} \right]^{-1/2} \left\{ g_{\alpha\beta} \frac{dx^\alpha}{dt} \frac{d\delta x^\beta}{dt} + \frac{1}{2} \frac{\partial g_{\alpha\beta}}{\partial x^\sigma} \delta x^\sigma \frac{dx^\alpha}{dt} \frac{dx^\beta}{dt} \right\} dt \,. \tag{1.95}$$

For simplicity we choose the parameterisation of $C(t)$ so that

$$g_{\mu\nu} T^\mu T^\nu = g_{\mu\nu} \frac{dx^\mu}{dt} \frac{dx^\nu}{dt} = 1 \,. \tag{1.96}$$

Then the first factor in (1.95) is unity, and the condition $\delta l = 0$ becomes, after integrating the first term in the sum by parts,

$$0 = \int_a^b \left\{ -\frac{d}{dt} \left(g_{\alpha\beta} \frac{dx^\alpha}{dt} \right) + \frac{1}{2} \frac{\partial g_{\alpha\lambda}}{\partial x^\beta} \frac{dx^\alpha}{dt} \frac{dx^\lambda}{dt} \right\} \delta x^\beta dt \,. \tag{1.97}$$

As the endpoints p and q are being kept fixed there is no contribution from the boundaries. This equation is satisfied for arbitrary δx^β only if

$$- g_{\alpha\beta} \frac{d^2 x^\alpha}{dt^2} - \frac{\partial g_{\alpha\beta}}{\partial x^\lambda} \frac{dx^\alpha}{dt} \frac{dx^\lambda}{dt} + \frac{1}{2} \frac{\partial g_{\alpha\lambda}}{\partial x^\beta} \frac{dx^\alpha}{dt} \frac{dx^\lambda}{dt} = 0 \tag{1.98}$$

Using the definition of the Christoffel symbols (1.68) we can see that this equation is precisely the geodesic equation (1.91). The same conclusion for a time-like curve can be reached with an analogous computation. We can conclude that the geodesic equation can be obtained by a variation of the Lagrangian[6]

$$\mathcal{L} = \sqrt{g_{\mu\nu} \frac{dx^\mu}{dt} \frac{dx^\nu}{dt}} \,. \tag{1.99}$$

[6] In this specific case the Lagrangian $\mathcal{L}' = g_{\mu\nu} \frac{dx^\mu}{dt} \frac{dx^\nu}{dt}$, for which the variational calculation is simpler, would work as well.

On a Riemannian manifold the length of curves connecting two points will be bounded from below, and the curve with that length (if it exists) will be a geodesic. However, a geodesic need not be the shortest path between two points (an example is furnished by the two arcs of the great-circle on a sphere connecting two points). For a Lorentzian manifold, a curve that maximises proper time is a geodesic (but again, a geodesic does not need to maximise proper time between two points).

Exponential map and normal coordinates

We can use the existence and uniqueness of geodesics to construct special coordinate systems. We start by defining the *exponential map* as the map from the tangent space T_p at a point $p \in \mathcal{M}$ to $\mathcal{M}$ obtained by mapping each tangent vector $\mathbf{T}$ into the point in $\mathcal{M}$ that lies along the geodesic through p with tangent $\mathbf{T}$ at a value $t = 1$ of the affine parameter. In general we will encounter obstacles like the crossing of geodesics or reaching a singularity, but one can show that there exists always a (sufficiently small) neighbourhood of p in which the exponential map exists and is one-to-one. More formally, for a geodesic γ_T defined by $\gamma_T(0) = p$ and $\dot{\gamma}_T(0) = \mathbf{T}$, where $\mathbf{T}$ is a tangent vector in a neighbourhood of 0 of T_p, the exponential map is defined as

$$\exp_p : \mathbf{T} \mapsto \gamma_T(1) . \tag{1.100}$$

As the tangent space is n-dimensional and can be identified with $\mathbb{R}^n$, we can then use the exponential map to provide us a with a coordinate system at p, called local inertial coordinates or *Riemann normal coordinates*. To do this, we start with an orthonormal basis $\{\mathbf{e}_\mu\}$ of T_p,

$$g_{\mu\nu} = \boldsymbol{g}(\mathbf{e}_\mu, \mathbf{e}_\mu) = \eta_{\mu\nu} . \tag{1.101}$$

We then use that for any point q close enough to p there is a unique geodesic that connects q to p, and a unique parametrisation t so that p is at $t = 0$ and q at $t = 1$. The tangent vector to this geodesic at p in the orthonormal basis is $\mathbf{k} = k^\mu \mathbf{e}_\mu$. We now assign to the point q the coordinates k^μ, i.e. we define the coordinates $x^\mu(q)$ as the components of the tangent vector that gets mapped to q by the exponential map, in an orthonormal basis.

In this coordinate system, geodesics through p are mapped into straight lines through the origin of $\mathbb{R}^n$ by construction: a set of vectors λk^μ for a fixed k^μ (a ray) in tangent space is mapped to a geodesic by the exponential map, and has coordinates $x^\mu(\lambda) = \lambda k^\mu$ in Riemann normal coordinates, satisfying therefore

$$\frac{d^2 x^\mu}{d\lambda^2} = 0 . \tag{1.102}$$

If all geodesics are straight lines then we can see from Eq. (1.91) that the Christoffel symbols vanish at p. This also implies that in Riemann normal coordinates, the covariant derivative at the origin is simply given by the usual partial derivative. With the metric compatibility $\nabla_\sigma g_{\mu\nu} = 0$ we then also have that $\partial_\sigma g_{\mu\nu} = 0$.

Geodesic deviation equation

We also derive the *geodesic deviation equation* which describes how the separation between two close geodesic curves changes. On the one hand, we will use the result very soon to motivate the field equation of General Relativity, and on the other hand the geodesic deviation equation describes for example how images of objects are distorted due to gravitational lensing, and hence provides a link between theoretical calculations in GR and observations. This is thus an important result.

We denote a one-parameter family of geodesics as $\gamma_s(t)$ so that for a fixed $s \in \mathbb{R}$ the curve $\gamma_s(t)$ is a geodesic with affine parameter t and where the map $(s,t) \to \gamma_s(t)$ is smooth, one-to-one and has a smooth inverse. We call the two-dimensional sub-manifold spanned by $\gamma_s(t)$ Σ, and we can choose s and t as coordinates on it. The vector field $T^\mu = (\partial/\partial t)^\mu$ is tangent to the geodesics and thus satisfies

$$T^\mu \nabla_\mu T^\nu = 0\,. \tag{1.103}$$

The vector field $X^\nu = (\partial/\partial s)^\nu$ describes the infinitesimal displacement between geodesics. It is called the deviation vector. Since T^μ and X^ν are coordinate vector fields, they commute, which from (1.64) also implies that

$$T^\mu \nabla_\mu X^\nu = X^\mu \nabla_\mu T^\nu\,. \tag{1.104}$$

The rate of change of the displacement vector between a geodesic and an infinitesimally close one, along the geodesic, is given by $V^\nu = T^\mu \nabla_\mu X^\nu$. We can then interpret

$$A^\nu = T^\mu \nabla_\mu V^\nu = T^\mu \nabla_\mu \left(T^\lambda \nabla_\lambda X^\nu \right) \tag{1.105}$$

as the relative acceleration of infinitesimally close geodesics in $\gamma_s(t)$. A short calculation shows the link between A^ν and curvature:

$$\begin{aligned} A^\nu &= T^\mu \nabla_\mu \left(T^\lambda \nabla_\lambda X^\nu \right) = T^\mu \nabla_\mu \left(X^\lambda \nabla_\lambda T^\nu \right) \\ &= \left(T^\mu \nabla_\mu X^\lambda \right) (\nabla_\lambda T^\nu) + X^\lambda T^\mu \nabla_\mu \nabla_\lambda T^\nu \\ &= \left(X^\mu \nabla_\mu T^\lambda \right) (\nabla_\lambda T^\nu) + X^\lambda T^\mu \nabla_\lambda \nabla_\mu T^\nu - R^\nu_{\mu\lambda\sigma} X^\lambda T^\mu T^\sigma \\ &= X^\mu \nabla_\mu \left(T^\lambda \nabla_\lambda T^\nu \right) - R^\nu_{\mu\lambda\sigma} X^\lambda T^\mu T^\sigma \\ &= -R^\nu_{\mu\lambda\sigma} X^\lambda T^\mu T^\sigma\,. \end{aligned} \tag{1.106}$$

This is known as the geodesic deviation equation. We can see that if the Riemann tensor is non-zero then some geodesics will accelerate towards or away from other geodesics and initially parallel geodesics ($V^\mu = 0$ initially) will fail to remain parallel. Since geodesics are our 'straight lines' (and are indeed straight lines if the Christoffel symbols are zero), we see explicitly that Euclid's fifth axiom (the parallel postulate) is indeed an independent axiom for flat space and is in general not true – and since we ascribe this failure of Euclid's fifth axiom to curvature, we also see the link between curvature and the Riemann tensor.

1.3 The Einstein equation

1.3.1 Galilean space and general covariance

We are now ready to return to the ideas in the first section and to apply them within the mathematical framework that we have developed in the second section. For example, we can see that the metric associated with pre-relativistic physics can be written as

$$ds^2 = (dx^1)^2 + (dx^2)^2 + (dx^3)^2 = h_{\mu\nu} dx^\mu dx^\nu \tag{1.107}$$

with $h_{\mu\nu} = \mathrm{diag}(1,1,1)$. We remind the reader that although we have expressed the metric above in a coordinate basis, the metric itself is a tensor field h that is independent of a choice of basis.

As the metric tensor h is constant in a Cartesian coordinate basis, the Christoffel symbols vanish in this coordinate system and thus the covariant derivative is just the normal derivative ∂_μ, and the curvature tensor vanishes. In other words, the metric $h_{\mu\nu}$ is flat. From the geodesic equation we see that the geodesics are straight lines in the sense that their Cartesian coordinates are linearly related to the affine parameter. The space of pre-relativity physics is thus the manifold $\mathbb{R}^3$ with a flat Riemannian metric.

An important principle in deriving physical equations is called *general covariance*. The principle says that the only quantity related to space or space-time that can show up in the description of physics is the metric. This should be seen in context with the tensorial nature of physical laws: We require that the laws of physics are independent of a specific choice of coordinate system, which implies that they should be expressible as tensor equations (with only few exceptions, e.g. generalisations like spinors). Tensors transform in a specific way under changes of the coordinates, precisely so that they represent geometric quantities that do not depend on the choice of coordinate system in which they are expressed.

A choice of preferred coordinate system can be seen as a choice of preferred vectors v^μ that form a basis of that coordinate system. However, general covariance states that such vectors are not permissible, as only the metric can be used. This implies for example also that the Christoffel symbols $\Gamma^\lambda_{\mu\nu}$, not being the components of a tensor, can not appear explicitly in a physical law, only in the form of the covariant derivative.

We should distinguish the concept of *invariance* from the concept of *covariance*[7]. Invariance means that any "absolute" objects of the theory are left invariant under the action of the appropriate group (see e.g. section 3.5 of [2] for a more detailed discussion). The metric $h_{\mu\nu}$ of Galilean space is such an absolute object, and hence the invariance group is given by the isometries, which is the group of Galilean transformations (rotations, translations and the discrete parity symmetry). To jump a bit ahead, the isometries (or invariance group) of Special Relativity is then the Poincaré or inhomogeneous Lorentz group that leaves the Minkowski metric $\eta_{\mu\nu}$ invariant, while in General Relativity, for which we will the demand the absence of any absolute elements, invariance and general covariance agree and the relevant group is the group of diffeomorphisms of $\mathbb{R}^4$. The name "General Relativity" appears thus justified.

1.3.2 Special Relativity

In Special Relativity we cannot split space and time and instead need to consider space-time, which has the manifold structure $\mathbb{R}^4$. The discussion in the first section motivates the definition of the metric of space-time

$$ds^2 = \eta_{\mu\nu} dx^\mu dx^\nu \tag{1.108}$$

with $\eta_{\mu\nu} = \mathrm{diag}(-1, 1, 1, 1)$, where x^μ is a global inertial coordinate system, and where the tensor field $\eta_{\mu\nu}$ is independent of the choice of global coordinate system. In this coordinate system the metric is again constant, so that the derivative operator associated to the metric is again the normal partial derivative ∂_μ. The manifold is once more flat, so that we can say that space-time is a flat Lorentzian $\mathbb{R}^4$ manifold. The geodesics of $\eta_{\mu\nu}$ are straight lines in a global inertial coordinate system – the time-like geodesics are the world-lines of inertial observers.

We can parametrise time-like curves with their proper time

$$\tau = \int \sqrt{-\eta_{\mu\nu} T^\mu T^\nu}\, dt \tag{1.109}$$

where t represents an arbitrary parameterisation of the curve, and where T^μ is the tangent to the curve in that parametrisation. In Special Relativity, τ is the time that would elapse on a clock carried along the curve (so different time-like curves connecting the same two events can correspond to different amounts of proper time, which is known as the twin-paradox). As discussed in the last section, the curve with the maximal amount of proper time is a geodesic (which corresponds to inertial motion).

[7] [1] uses "special covariance" instead of "invariance".

The tangent vector u^μ to a time-like curve parameterised by τ is called the 4-velocity of the curve, because of (1.109) it satisfies

$$u^\mu u_\mu = -1\,. \tag{1.110}$$

An inertial observer, which corresponds to an observer not subject to any external forces, travels on a geodesic. The 4-velocity of an inertial observer thus satisfies the equation of motion

$$u^\mu \partial_\mu u^\nu = 0\,. \tag{1.111}$$

The momentum 4-vector of a particle is defined as

$$p^\mu = m u^\mu \tag{1.112}$$

where m is the rest-mass of the particle. The energy of the particle as measured by an observer with 4-velocity v^μ is

$$E = -p_\mu v^\mu\,. \tag{1.113}$$

For an observer at rest in the coordinate system the 4-velocity is $v^\mu = (1,0,0,0)$ and thus the energy can be considered as the time-component of the momentum 4-vector. For an observer for whom the particle is at rest, $v^\mu = u^\mu$, the energy is given by the familiar formula $E = mc^2$, where we usually set $c = 1$.

1.3.3 The stress - energy tensor in Special Relativity

Continuous matter distributions (*fluids*) are described in Special Relativity with a symmetric rank-2 tensor $T_{\mu\nu}$, called *energy-momentum* or *stress-energy* tensor. For an observer with 4-velocity v^μ the scalar $T_{\mu\nu}v^\mu v^\nu$ is interpreted as the energy density as measured by this observer. For normal matter, the energy density is non-negative for all observers, $T_{\mu\nu}v^\mu v^\nu \geq 0$. The energy density is equal to the component T_{00} for an observer at rest in that coordinate system.

If x^μ is orthogonal to v^μ then $-T_{\mu\nu}v^\mu x^\nu$ is interpreted as the momentum density of the matter in the x^μ direction. If y^μ is also orthogonal to v^μ then $-T_{\mu\nu}x^\mu y^\nu$ is the $x^\mu - y^\nu$ component of the stress tensor of the material. This element of $T_{\mu\nu}$ corresponds to the y^ν component of the force on a unit area perpendicular to the x^μ direction. The stress tensor is symmetric in order to avoid inducing a non-vanishing torque on a volume element even for a spatially constant stress tensor.

A *perfect fluid* is defined as a continuous distribution of matter with stress-energy tensor of the form

$$T_{\mu\nu} = \rho u_\mu u_\nu + P\left(\eta_{\mu\nu} + u_\mu u_\nu\right)\,, \tag{1.114}$$

where u_μ is a time-like unit vector field representing the 4-velocity of the fluid. We see that ρ and P are therefore the mass - energy density and the pressure respectively, as measured in the fluid rest-frame. The fluid is called “perfect” as it has no viscosity – in the fluid rest-frame the stress tensor is diagonal, meaning that no shear stresses that act tangentially on a surface element exist, only a uniform pressure acting in the normal direction.

The equation of motion of a perfect fluid, in the absence of external forces, is simply

$$\partial^\mu T_{\mu\nu} = 0\,. \tag{1.115}$$

We can project this equation along u^μ (the ‘time’ component) and perpendicular to it (the ‘space’ components, using the orthogonal projector[8] $P_\mu^\nu = \delta_\mu^\nu + u_\mu u^\nu$), giving us in terms of ρ, P and u^μ

$$u^\mu \partial_\mu \rho + (\rho + P)\partial^\mu u_\mu = 0\,, \tag{1.116}$$

$$(\rho + P) u^\mu \partial_\mu u_\nu + \left(\eta_{\mu\nu} + u_\mu u_\nu\right)\partial^\mu P = 0\,. \tag{1.117}$$

[8] P_μ^ν is a projector because $P_\mu^\lambda P_\lambda^\nu = P_\mu^\nu$ and it projects onto the orthogonal subspace to u^μ since $P_\mu^\nu u_\nu = 0$.

In the non-relativistic limit we have that $P \ll \rho$ (as ρ is dominated by the mass energy), $u^\mu = (1, \mathbf{v})$ and $v dP/dt \ll |\nabla P|$ (where we neglected factors of c as we use $c = 1$), and the equations simplify to

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0 \,, \tag{1.118}$$

$$\rho \left\{ \frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla)\mathbf{v} \right\} = -\nabla P \,. \tag{1.119}$$

The first equation is the conservation of mass, and the second equation expresses Newton's second law (balance of momentum) for a perfect fluid, often called Euler's equation. Both are contained in Eq. (1.115). We can consider that equation also from a more general angle: Let v^μ be a group of inertial observers with parallel four-velocities, $\partial_\mu v^\nu = 0$. The quantity

$$J_\mu = -T_{\mu\nu} v^\nu \tag{1.120}$$

is the mass-energy current density 4-vector of the fluid as measured by these observers. Equation (1.115) then implies

$$\partial^\mu J_\mu = 0 \,. \tag{1.121}$$

The fact that J_μ is divergence free corresponds, through Gauss' law, to a conservation law for energy: Integrating J_μ over the boundary ∂V of a four-dimensional space-time volume V we have that

$$\int_{\partial V} J_\mu n^\mu dS = 0 \tag{1.122}$$

where n^μ is the unit normal vector to the surface and dS the surface volume element. This equation says that the energy change in a spatial volume (the spatial bottom and top 'caps' of V) equals the energy flux into and out of the volume, integrated over time (the 'sides' of V). Thus, if we want inertial observers to find that energy is conserved we need (1.115) to hold generally (not only for a perfect fluid). Any theory of relativity should therefore not contradict this property, a point that will become important soon.

1.3.4 General Relativity

We already discussed in the introduction that Maxwell's theory of electromagnetism fits beautifully into the framework of Special Relativity, and that it would have seemed natural to look for ways to extend Newton's theory of gravity in a similar way. However, we also discussed that the Principle of Equivalence places strong obstacles in the way of such a plan.

In addition to the gravitational redshift discussed earlier, we can also see it as follows: When we measure the electromagnetic field, we start by setting up neutral inertial observers that do not interact with the electromagnetic field. We then add a test charge and measure how the trajectory of a test charge deviates from those of the inertial observers. In this way we can map out the electromagnetic field. However, this approach is difficult to implement for a gravitational force field, because of the equivalence principle. Since everything is subject to gravity, one would need to postulate imaginary 'non-gravitational' inertial observers that move according to the flat Special Relativistic space-time metric, and then consider the motion of test particles with respect to those imaginary observers to map out a 'gravitational field'. In such a set-up, the equivalence principle does not follow naturally, it is imposed by hand. Einstein instead chose to change the point of view, incorporating the principle of equivalence as a fundamental principle into his theory, by postulating that it is the metric of space-time itself that changes, and that there is no gravitational force field (and no generally meaningful way to define one).[9] Instead inertial observers fall freely along geodesics of the metric.

[9] Note that for macroscopic bodies the relative gravitational force (tidal field) of gravity between different points is also measurable in General Relativity, for example through the geodesic deviation equation (1.106). Indeed, as this equation shows the relative acceleration is directly linked to the curvature of space-time.

In this context, the apparent existence of a gravitational force field at the surface of the Earth arises precisely due to the presence of the surface, which prevents us from following geodesics of the metric and instead accelerates us 'upward'. Because the surface of the Earth is to a good approximation (on human time scales) static, we effectively appear to be at rest in spite of the acceleration. In this specific situation we can use this property (time translation symmetry provided by the static surface of the Earth) to set up preferred observers, with respect to whom we measure gravity – indeed, that is what we do in our mind automatically. This is however a special situation, in general there is no such time translation symmetry that keeps accelerated observers 'at rest', and thus there is no set of preferred non-inertial observers that can be used to define a gravitational force.

Since in Einstein's theory the metric is no longer the flat metric $\eta_{\mu\nu}$ of Special Relativity, but instead a general metric $g_{\mu\nu}$ of a curved space-time, is appears then also natural to consider a general manifold $\mathcal{M}$ rather than the $\mathbb{R}^4$ of Special Relativity. Thus our generalized theory of gravity should be based on a space-time that is a general manifold $\mathcal{M}$ with a Lorentzian metric $g_{\mu\nu}$. But since Special Relativity works rather well for Electrodynamics, it seems reasonable to extend it in a minimal way. Of course the resulting theory of gravity needs to be confronted with observations to test whether it is acceptable.

General Relativity is governed by two basic principles:

1. The law of general covariance, which states that the metric $g_{\mu\nu}$, and (tensorial) quantities derived from it, are the only space-time quantities allowed in equations of physics;

2. The requirement that equations must reduce to the equations satisfied in Special Relativity in the case where $g_{\mu\nu}$ is flat.

These two principles will serve as guides in the construction of the new theory[10]. Effectively we will be enlarging Special Relativity with the extension of flat $\mathbb{R}^4$ to a general manifold with a Lorentzian metric. However, the physical content of the theory will still be described by the same tensor fields as before. Thus, in GR particle motion is still described by a time-like curve, perfect fluids by a 4-velocity u^μ, density ρ and pressure P, and the electromagnetic field by an antisymmetric field-strength tensor $F_{\mu\nu}$. Only the equations satisfied by these quantities need to be amended, in a way that respects the two principles above. The minimal extension for any equation holding in Special Relativity is the following:

- Replace $\eta_{\mu\nu}$ with $g_{\mu\nu}$.

- Replace the derivative ∂_μ with the covariant derivative ∇_μ associated with $g_{\mu\nu}$.

The second rule is related to minimal coupling in electromagnetism (or more generally in gauge theories) where the derivative ∂_μ is replaced with the gauge-covariant derivative $D_\mu = \partial_\mu - ieA_\mu$. We should however mention that the application of these rules is not always unique, so that occasionally care must be taken in order to get the right result (we will see an example later on, in Section 1.4.2).

Based on these rules, the 4-velocity of a particle u^μ is still the unit tangent (now based on $g_{\mu\nu}$) to its world-line, as in Special Relativity. A free particle follows the geodesic equation of motion

$$u^\mu \nabla_\mu u^\nu = 0\,. \tag{1.123}$$

We can define the acceleration of a particle $a^\nu = u^\mu \nabla_\mu u^\nu$ (not the same as the relative acceleration of geodesics in the geodesic deviation equation), and if it is non-vanishing then we say that a force $f^\mu = ma^\mu$ acts on the particle, where m is the rest-mass. For example, for a particle with mass m and charge q the

[10] We will add later that in the limit of weak fields and non-relativistic motion we should find Newtonian gravity to a good approximation.

equation of motion in the presence of an electromagnetic field $F_{\mu\nu}$, taking into account the resulting Lorentz force, is

$$u^\mu \nabla_\mu u^\nu = \frac{q}{m} F^\nu{}_\rho u^\rho \,, \tag{1.124}$$

where we now of course raise and lower indices (e.g. on $F_{\mu\nu}$) with the metric $g_{\mu\nu}$. The 4-momentum of a particle is defined by

$$p^\mu = m u^\mu \,, \tag{1.125}$$

and the energy of the particle, as measured by an observer with 4-velocity v^μ present on the particles world-line, is

$$E = -p_\mu v^\mu \,. \tag{1.126}$$

All of these equations are directly 'translated' from Special Relativity to General Relativity. A difference is however that now there is no general notion of vectors being parallel, since on a curved manifold parallel transport is path dependent. For this reason there is also no natural global set of inertial observers, and a given observer cannot, in general, define the energy of a distant particle.

As in Special Relativity, continuous matter distributions are also in General Relativity described by a stress-energy tensor $T_{\mu\nu}$. For a perfect fluid it is given by

$$T_{\mu\nu} = \rho u_\mu u_\nu + P(g_{\mu\nu} + u_\mu u_\nu) \tag{1.127}$$

which satisfies now the covariant conservation equation

$$\nabla^\mu T_{\mu\nu} = 0 \,. \tag{1.128}$$

This equation gives, as for Eq. (1.115),

$$u^\mu \nabla_\mu \rho + (\rho + P)\nabla^\mu u_\mu = 0 \,, \tag{1.129}$$

$$(P + \rho) u^\mu \nabla_\mu u_\nu + (g_{\mu\nu} + u_\mu u_\nu)\nabla^\mu P = 0 \,. \tag{1.130}$$

Although (1.128) is usually still considered as a conservation equation, that interpretation is less clear in GR. For a family of observers represented by a time-like unit vector field v^μ that is covariantly constant ($\nabla_\mu v_\nu = 0$) or that satisfies at least $\nabla_{(\mu} v_{\nu)} = 0$ we would have $\nabla^\mu (T_{\mu\nu} v^\nu) = 0$. Gauss' law in curved space-time would then still give strict energy conservation for the energy-momentum 4-vector $J_\mu = -T_{\mu\nu} v^\nu$ measured by the observers v^μ. But in curved space-time one can in general no longer find such a v^μ that satisfies $v^\mu v_\mu = -1$ and $\nabla_{(\mu} v_{\nu)} = 0$[11]. Only in special situations or for small volumes (compared to the local curvature radius), where tidal forces can be neglected, is the energy of a fluid approximately conserved, and in such a small region we can also find vector fields with $\nabla_\mu v_\mu \approx 0$. So Eq. (1.128) provides a local conservation of mass-energy in small space-time volumes, and in this capacity it should still hold for all kinds of matter and fields, not only for perfect fluids.

The equation that does not exist in Special Relativity is the one that determines the metric. It is not necessary since the special relativistic metric is always given by $\eta_{\mu\nu}$. But now $g_{\mu\nu}$ is dynamical and it needs to be determined by some equation of motion. Motivated by Mach's principle, in General Relativity the geometry of space-time should be influenced by the matter and energy present in the Universe.

Since Newtonian gravity works quite well in a wide range of situations, it may be worthwhile to look also there for inspiration. We would like to find a generally covariant equation that agrees with the Newtonian theory of gravity in the limit of weak fields and slowly moving matter. But what quantity should we consider? We have seen in Section 1.2.7 that locally we can transform away the Christoffel symbols by using Riemann normal coordinates. The curvature, being a tensor, cannot be fully transformed

[11]This is actually Killing's equation, see Section 3.1.3

away as the zero tensor would be invariant under coordinate transforms. As we computed in Eq. (1.106), the curvature determines the relative acceleration of nearby freely falling test-bodies, which corresponds to the "tidal forces" acting on them.

In Newtonian theory, the gravitational field is given by a potential ϕ, and the equations of motion of two nearby particles, separated by a vector $\mathbf{d}$, are given by

$$\ddot{x}^i(t) = -\left.\frac{\partial\phi}{\partial x^i}\right|_{\mathbf{x}(t)} \tag{1.131}$$

$$\ddot{x}^i(t) + \ddot{d}^i(t) = -\left.\frac{\partial\phi}{\partial x^i}\right|_{\mathbf{x}(t)+\mathbf{d}(t)} = -\left.\frac{\partial\phi}{\partial x^i}\right|_{\mathbf{x}(t)} - \left.\frac{\partial^2\phi}{\partial x^i\partial x^j}\right|_{\mathbf{x}(t)} d^j(t) + \mathcal{O}(d^2)\,. \tag{1.132}$$

Taking the difference of the two equations, we find that to lowest order in $\mathbf{d}$ the tidal (relative) acceleration of the particles is given by

$$\ddot{d}^i(t) = -\frac{\partial^2\phi}{\partial x^i\partial x^j} d^j(t)\,. \tag{1.133}$$

In GR on the other hand, the tidal acceleration is given by $-R^{\nu}_{\mu\lambda\sigma}u^\mu d^\lambda u^\sigma$ from (1.106), where u^μ is the 4-velocity of the particles, and d^μ the deviation vector. This suggests a correspondence

$$R^{\nu}_{\mu\lambda\sigma}u^\mu u^\sigma \leftrightarrow \partial^\nu\partial_\lambda\phi\,. \tag{1.134}$$

In addition, the Newtonian gravitational potential is determined through the Poisson equation (for $G = c = 1$)

$$\nabla^2\phi = 4\pi\rho\,. \tag{1.135}$$

At the beginning of Section 1.3.3 we argued that the energy density in SR is described through $T_{\mu\nu}u^\mu u^\nu$ where $T_{\mu\nu}$ is the energy momentum tensor and u^μ the observer 4-velocity. This equation should still hold in GR, suggesting

$$T_{\mu\nu}u^\mu u^\nu \leftrightarrow \rho\,. \tag{1.136}$$

The Newtonian 'equation of motion' for gravity is the Poisson equation above. Inserting the correspondences naively into that equation, we obtain a relationship between the curvature and the stress-energy tensor, $R^{\nu}_{\mu\nu\sigma}u^\mu u^\sigma \sim \nabla^2\phi \sim \rho \sim T_{\mu\nu}u^\mu u^\nu$. The obvious proposal for GR then is $R_{\mu\nu} = 4\pi T_{\mu\nu}$, which is indeed the equation originally postulated by Einstein. However, as Einstein quickly realised, this equation has a serious issue: the stress-energy tensor satisfies the conservation equation $\nabla^\mu T_{\mu\nu} = 0$. We have seen that the Einstein tensor satisfies the equivalent equation $\nabla^\mu G_{\mu\nu} = 0$ thanks to the contracted Bianchi identity (1.88). Hence if indeed $R_{\mu\nu} = 4\pi T_{\mu\nu}$ then we would conclude that $\nabla^\mu R = 0$, meaning that R, and thus $T = T^\mu_\mu$ ($= 3P - \rho$ for a perfect fluid) would have to be constant throughout the universe, which is clearly not a physically acceptable constraint on the matter distribution. On the other hand, the *Einstein equation*

$$G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = 8\pi T_{\mu\nu} \tag{1.137}$$

does not have this problem, and it is also compatible with the Newtonian limit: The trace of the Einstein equation gives $R = -8\pi T$, so that we can also write

$$R_{\mu\nu} = 8\pi\left(T_{\mu\nu} - \frac{1}{2}g_{\mu\nu}T\right)\,. \tag{1.138}$$

In Newtonian theory the pressure does not gravitate so that $T \approx -\rho$ and hence the desired relation is preserved. The Einstein equation (1.137) is indeed the field equation of General Relativity that Einstein published in 1915 and that seems to provide an excellent description of gravity.

We can summarize General Relativity as follows:

Space-time is a manifold $\mathcal{M}$ with a Lorentzian metric $g_{\mu\nu}$ and associated Levi-Civita connection. The curvature of space-time is related to the matter distribution through Einstein's equation (1.137).

We note that the contracted Bianchi identity now *implies* $\nabla^\mu T_{\mu\nu} = 0$, and so provides an equation which describes the evolution of matter. For a perfect fluid, this equation gives a complete description of the evolution of the matter. For the case of $P = 0$, corresponding to collisonless dust, i.e. particles that exert no forces on each other, the fluid equation of motion (1.130), $u^\mu \nabla_\mu u_\nu = 0$, implies that individual dust particles move on geodesics. This result can be generalized, and it can be shown that $\nabla^\mu T_{\mu\nu} = 0$ implies that any sufficiently "small" body, whose self-gravity is sufficiently weak, must travel on geodesics. Therefore the geodesic hypothesis that we used earlier, namely that the world-lines of test particles are geodesics of the space-time metric, is already contained in Einstein's equation (for objects "large" enough the tidal forces of the gravitational field lead in general to deviations from geodesic motion, but it can be shown that also the equations of motion of such bodies can be found from $\nabla^\mu T_{\mu\nu} = 0$). Of course one can argue that we constructed the theory so that the covariant conservation of the stress-energy tensor is a consequence, but it nonetheless reveals an important consistency property of the Einstein equation in the framework of General Relativity.

From the discussion presented here, the development of General Relativity appears fairly straightforward. We should however not forget that it took Einstein about ten years to build his theory. We should also not forget that good theoretical motivations do not make a true theory in physics, nature always has the final word. Experimental tests are thus crucial, this is one of the motivations behind the study of solutions of General Relativity.

1.3.5 Solving the Einstein equations

The form of the Einstein equations looks like the stress-energy tensor acting as a source for the curvature of space-time. However, unlike e.g. Maxwell's equation we cannot, in general, just specify $T_{\mu\nu}$ and solve for $g_{\mu\nu}$, since without knowing the metric, we do not know how to interpret $T_{\mu\nu}$ physically – for example, Equation (1.127) for a perfect fluid contains the metric explicitly, and the same is true for other situations. For this reason it is necessary to solve simultaneously for the space-time metric and matter distribution, which makes it much harder to find consistent solutions of Einstein's equation. Exact solutions are only known in special, highly symmetric cases, some important examples are:

- The *Schwarzschild* metric: A static and spherically symmetric solution, it describes irrotational and uncharged black holes, and can be used as an approximation for spherical stars. We will discuss it in Chapter 3.

- The *Kerr* metric: A stationary and axisymmetric solution, generalizing the Schwarzschild solution to the case of a rotating black hole.

- The *Minkowski* space: Flat Minkowski space, $g_{\mu\nu} = \eta_{\mu\nu}$, is a solution of Einstein's equation for vacuum, $T_{\mu\nu} = 0$ (but not the only one, the solution depends on the boundary conditions!).

- The *de Sitter* space-time (and *Anti - de Sitter* space-time): The remaining maximally symmetric space-time solutions (together with Minkowski).

- The *Friedmann-Lemaître-Robertson-Walker* (FLRW) metric: A homogeneous and isotropic solution that is thought to describe the average behaviour of the Universe.

We will spend most of the rest of the GR lectures on discussing some of these.

The curvature tensor (of which the Ricci tensor and scalar are contractions, i.e. linear sums) consists of first derivatives and squares of Christoffel symbols, cf. Eq. (1.71). The Christoffel symbols in turn are built from a sum of first derivatives of the metric, (1.68). Mathematically, the Einstein equation can be expressed, in a coordinate basis, as a set of coupled non-linear partial differential equations (PDE) of the metric. The equations contain at most second derivatives of $g_{\mu\nu}$ and are linear in the second derivatives. For a Lorentzian metric the equations are hyperbolic, like a wave equation – we will see later in Chapter 2 that they describe, among other things, gravitational waves propagating in vacuum.

Is the Einstein equation enough to uniquely determine a solution? A solution would be uniquely determined if the metric tensor field is fully known, as we can compute e.g. the Riemann tensor from it. We can easily see that, given a stress-energy tensor $T_{\mu\nu}$ we can determine the Ricci tensor but not the full Riemann tensor as we have no information on the Weyl part. As an example, there are many possible vacuum solutions, i.e. solutions for $T_{\mu\nu} = 0$, of Einstein's equation. This is however not a surprise when we consider the Einstein equation as a system of coupled PDEs for the metric, as in this case we know that we have to include initial and boundary values in order to determine uniquely the solution.

The question then becomes the so-called "Cauchy problem" of GR, namely if $\mathcal{S}$ is a three-dimensional Riemannian manifold with metric γ_{ij} (the "initial spatial hypersurface"), and if $\gamma_{0\mu}$ and $\partial_0\gamma_{\mu\nu}$ are given functions (all smooth), is there a Lorentz manifold $(\mathcal{M}, g)$ and an embedding $\sigma : \mathcal{S} \to \mathcal{M}$ such that, on $\sigma(\mathcal{S})$, $g_{\mu\nu}$ and its time derivative agree with $\gamma_{\mu\nu}$ and $\partial_0\gamma_{\mu\nu}$ and so that

$$G_{\mu\nu} = 8\pi T_{\mu\nu} \tag{1.139}$$

on $\mathcal{M}$? In general this is not a trivial question, a more detailed discussion can be found for example in Section 3.8 of [2]. Here we will just show that in principle it should work.

The metric, being a 4×4 symmetric tensor field, contains 10 degrees of freedom. The Einstein equation provides also 10 PDE's for $g_{\mu\nu}$ – but an important point that we have left out so far is that because of the Bianchi identity, $\nabla_\nu G^{\mu\nu} = 0$, which effectively removes four equations. We can think of these four equations as "constraint equations", as the Bianchi identity gives

$$G^{\mu 0}_{\ ,0} = -G^{\mu i}_{\ ,i} - \Gamma^{\mu}_{\nu\lambda} G^{\lambda\nu} - \Gamma^{\nu}_{\nu\lambda} G^{\mu\lambda} \,. \tag{1.140}$$

The right hand side contains at most second derivatives wrt time, so that $G^{\mu 0}$ contain at most first derivatives of time. For this reason the equations

$$G_{\mu 0} = 8\pi T_{\mu 0} \tag{1.141}$$

place constraints on $g_{\mu\nu}$ and $g_{\mu\nu,0}$ that have to be satisfied on $\sigma(\mathcal{S})$ for a solution to exist.

The remaining six "evolution equations"

$$G_{ij} = 8\pi T_{ij} \tag{1.142}$$

then determine the evolution of the ten components of $g_{\mu\nu}$ as a function of time. The solution is thus not uniquely determined – but that *should* be the case, as we have the freedom to choose coordinates as we wish, which corresponds precisely to the "missing" four degrees of freedom.

These four degrees of freedom can be removed from the problem by imposing gauge conditions, an example is the *harmonic gauge*, defined by the four conditions

$$\left(\sqrt{-g} g^{\mu\nu}\right)_{,\nu} = 0 \,, \tag{1.143}$$

where $g = \det(g_{\mu\nu})$. The time derivative of these conditions gives

$$\left(\sqrt{-g} g^{\mu 0}\right)_{,00} = \left(\sqrt{-g} g^{\mu i}\right)_{,0i} \,. \tag{1.144}$$

These equations together with the six equations $G_{ij} = 8\pi T_{ij}$ then provide ten second-order in time differential equations for the ten functions $g_{\mu\nu}$.

For initial conditions that satisfy the constraints (1.141), a gauge choice that is compatible with the problem and a "reasonable" stress-energy tensor one can show that it is always possible to find a local solution of (1.142). The existence of global solutions is much more difficult to prove, and general results are limited.

1.4 Lagrangian formulation of GR and the Einstein-Hilbert action

It is possible to write down a Lagrangian density for General Relativity, which then provides an action principle on which GR can be based. This is useful in many contexts, not least for studying systematically extensions to GR. However, as we will see, it also provides a way to compute the stress-energy tensor from the action of any "matter" fields that are present.

1.4.1 The Einstein-Hilbert action

A powerful technique in mechanics and field theory is the use of Lagrangian mechanics and the principle of least (or, more generally, stationary) action. For example, the Lagrangian density of the massive scalar field is given by

$$\mathcal{L}_\phi = -\frac{1}{2}\left(\partial_\mu\phi\partial^\mu\phi + m^2\phi^2\right) , \tag{1.145}$$

and the action of the scalar field in special relativity is then

$$S_\phi = \int_\mathcal{M} \mathcal{L}_\phi \, . \tag{1.146}$$

In SR, the integration is over the Minkowski space-time. From here we can use a variational principle, leading to the Euler-Lagrange equation, to compute the equation of motion of ϕ.

In GR, the Lagrangian formulation has one main technical difficulty, namely the volume element that should be used in integrals like (1.146). We skip here over the formal definition of integration on manifolds, discussions can be found for example in Appendix B of [1] or Sections 14.6.1 and 14.7.2 of [2]. Instead we motivate the volume element simply from the requirement that an integral of the type

$$\int_\mathcal{M} f d\mu_g \tag{1.147}$$

of a function f defined on a manifold $\mathcal{M}$, for a measure $d\mu_g$ that may depend on the metric, should be independent of a choice of coordinates, and hence be invariant under smooth coordinate changes $\{x^i\} \to \{\tilde{x}^i\}$. As the determinant of a product of matrices is the product of the determinants, we have for the determinant of the metric tensor

$$\sqrt{|\tilde{g}(\tilde{x})|} = \sqrt{|g(x)|}\left|\det\left(\frac{\partial x^k}{\partial \tilde{x}^l}\right)\right| \, . \tag{1.148}$$

From this and the transformation property of integrals it follows that for a continuous function f with compact support in a coordinate patch $U \subset \mathcal{M}$ the integral

$$\int_\mathcal{M} f d\mu_g \equiv \int_U f(x)\sqrt{|g(x)|} d^4x \tag{1.149}$$

is independent of the coordinate system and has the correct flat-space limit (in which case $|g(x)| = |\eta| = 1$). If the support of f is not in the domain of any single chart then we need to use a partition of unity (a selection of weight functions ψ_i so that every ψ_i has only support in a patch U_i and that verifies $\sum_i \psi_i = 1$ everywhere) to combine several charts, in order to define $d\mu_g$. Often $d\mu_g$ is called the measure associated to the metric g. For a Lorentzian metric, one usually writes $\sqrt{-g}$ instead of the absolute value of the determinant of the metric.

We see that the volume element now depends explicitly on the metric, and hence it needs to be included in the variation if the metric itself is the field variable in the variational principle. In GR therefore we should use $\sqrt{-g}\mathcal{L}$ for the Lagrangian density.

We now show that, except potentially for boundary terms, $\mathcal{L}_{\text{EH}} = R$ is a Lagrangian density for the vacuum Einstein equations[12]. The corresponding action

$$S_{\text{EH}} = \int R d\mu_g = \int \sqrt{-g}\, R\, d^4x \tag{1.150}$$

is called the *Einstein-Hilbert* action. Here we assume that the integral is over a compact region D with smooth boundary ∂D. Then the first variation of (1.150) is given by

$$\delta S_{\text{EH}} = \int_D G_{\mu\nu}\delta g^{\mu\nu} d\mu + \int_D \text{div} W d\mu \tag{1.151}$$

where W is a vector field that will be defined later. We show this by explicitly computing the variation of the Einstein-Hilbert action,

$$\delta \int_D d\mu R = \int_D \delta\left(g^{\mu\nu} R_{\mu\nu}\sqrt{-g}\right) d^4x = \int_D \delta R_{\mu\nu} g^{\mu\nu}\sqrt{-g} d^4x + \int_D R_{\mu\nu}\delta\left(g^{\mu\nu}\sqrt{-g}\right) d^4x\,. \tag{1.152}$$

To compute $\delta R_{\mu\nu}$ we use (1.71), which implies

$$R_{\mu\nu} = \partial_\rho \Gamma^\rho_{\mu\nu} - \partial_\nu \Gamma^\rho_{\mu\rho} + \Gamma^\sigma_{\mu\nu}\Gamma^\rho_{\sigma\rho} - \Gamma^\sigma_{\mu\rho}\Gamma^\rho_{\sigma\nu}\,. \tag{1.153}$$

It is easiest to now use Riemann normal coordinates, cf Section 1.2.7, with the origin at the point $p \in D$ (but the calculation is the same without this trick, it is just a bit longer). Then at p only the derivatives of the Christoffel symbols are non-zero and we get

$$\delta R_{\mu\nu} = \partial_\rho \delta\Gamma^\rho_{\mu\nu} - \partial_\nu \delta\Gamma^\rho_{\mu\rho}\,. \tag{1.154}$$

From the transformation law for Christoffel symbols (1.51) we see that the variations $\delta\Gamma$ will transform as tensors as the variation of the constant part vanishes. At p (where the Christoffel symbols vanish) we can then write

$$\delta R_{\mu\nu} = \nabla_\rho \delta\Gamma^\rho_{\mu\nu} - \nabla_\nu \delta\Gamma^\rho_{\mu\rho}\,. \tag{1.155}$$

This is a tensor equation (known as *Palatini identity*) and for this reason holds everywhere (not only at p) and in all coordinate systems. (We could also have computed everything without using the Riemann normal coordinates and would have found the same result.) It implies that

$$g^{\mu\nu}\delta R_{\mu\nu} = \nabla_\rho W^\rho \tag{1.156}$$

for the vector field

$$W^\rho = g^{\mu\nu}\delta\Gamma^\rho_{\mu\nu} - g^{\rho\nu}\delta\Gamma^\mu_{\mu\nu}\,. \tag{1.157}$$

Thus, the first term in the last equation of (1.152) is the volume integral of $\text{div}W$.

For the second term we need to compute $\delta(\sqrt{-g})$, which you will do in the exercises. You should find

$$\delta\left(\sqrt{-g}\right) = -\frac{1}{2}\sqrt{-g} g_{\mu\nu}\delta g^{\mu\nu}\,. \tag{1.158}$$

The second term then gives

$$\int_D \left(R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu}\right)\delta g^{\mu\nu}\sqrt{-g}\, d^4x \tag{1.159}$$

[12] We could add a total divergence without changing the result. For example the Gauss-Bonnet scalar is in four dimensions locally a total divergence.

which is indeed the first term of (1.151). If the variations $\delta g^{\mu\nu}$ vanish outside a region contained in D then the integral over the total divergence, which only contributes a boundary term, is zero and only the first term in (1.151) survives. We then have that

$$\delta \int_D R d\mu = \int_D G_{\mu\nu} \delta g^{\mu\nu} d\mu \,. \tag{1.160}$$

Here we will always assume that this is the case. Otherwise we obtain a boundary term, that could be canceled by the appropriate modification of the Einstein-Hilbert action[13]. The variation of the Einstein-Hilbert action therefore results in the vacuum Einstein equations, $G_{\mu\nu} = 0$.

For later use we note that, based on the variation of $\sqrt{-g}$,

$$\delta(d\mu) = -\frac{1}{2} g_{\mu\nu} \delta g^{\mu\nu} d\mu \,. \tag{1.161}$$

1.4.2 Scalar and electromagnetic fields in General Relativity

The discussion so far covers the vacuum Einstein equations, i.e. the ‘gravity’ part of a Lagrangian, appropriate when there is no ‘matter’ present. In general we should add to the gravity Lagrangian also the matter Lagrangian. Before we consider such a combination, we study two important examples on their own, namely the scalar field and the electromagnetic field. The electromagnetic field is obviously an important quantity, and well adapted for inclusion in GR as it is already in the correct form for Special Relativity. Another important quantity is the scalar field. Although we have not found many fundamental scalar fields in nature yet (only the Higgs field), it is a common toy model for theorists, and a common effective description in many situations. Apart from their importance for describing physics (e.g. light rays), these fields also serve as a ‘worked example’ for how to go from Special Relativity to General Relativity, and we will encounter an explicit example where the ‘minimal extension’ approach to go from SR to GR requires some care.

Scalar fields in special and general relativity

A classical, massive scalar field ϕ satisfies in SR the *Klein-Gordon* equation

$$\partial^\mu \partial_\mu \phi - m^2 \phi = 0 \,. \tag{1.162}$$

Its stress-energy tensor is given by

$$T_{\mu\nu} = \partial_\mu \phi \partial_\nu \phi - \frac{1}{2} \eta_{\mu\nu} \left(\partial^\lambda \phi \partial_\lambda \phi + m^2 \phi^2 \right) \,. \tag{1.163}$$

In the exercises you will compute this tensor from the Lagrangian density of the massive scalar field, which is given by

$$\mathcal{L}_\phi = -\frac{1}{2} \left(\partial_\mu \phi \partial^\mu \phi + m^2 \phi^2 \right) \,, \tag{1.164}$$

and the action of the scalar field in special relativity is then

$$S_\phi = \int_{\mathcal{M}} \mathcal{L}_\phi \,. \tag{1.165}$$

Here the integration is over the Minkowski space-time, but in general we only consider variations within a compact region U of $\mathcal{M}$ that vanish on the boundary. It is easy to see that the Klein-Gordon equation

[13] Geometrically the boundary term is the variation of the trace of the extrinsic curvature of the boundary. Since we have not introduced the extrinsic curvature we skip this aspect here, in any case the boundary term is usually neglected.

(1.162) can be obtained from $\delta S_\phi/\delta\phi = 0$, using the Euler-Lagrange equation (generalized to the case of multiple variables),

$$\frac{\partial \mathcal{L}}{\partial \phi} - \frac{\partial}{\partial x^\mu}\frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} = 0\,. \tag{1.166}$$

As we will see in the next section in the context of GR, the stress-energy tensor can formally be obtained from the variation of the action with respect to the metric. While in GR this procedure makes sense, it is less clear why this should work in SR as there the metric is constant. An alternative approach that works (for scalar fields) also in SR, based on Noether's theorem, is discussed at the end of appendix E.1 of [1].

The most straightforward conversion to GR of the scalar field equation of motion (1.162) follows the 'minimal replacement' rule, here $\partial_\mu \to \nabla_\mu$,

$$\nabla^\mu \nabla_\mu \phi - m^2\phi = 0\,. \tag{1.167}$$

Correspondingly, the stress-energy tensor (1.163) becomes

$$T_{\mu\nu} = \nabla_\mu\phi\nabla_\nu\phi - \frac{1}{2}g_{\mu\nu}\left(\nabla^\lambda\phi\nabla_\lambda\phi + m^2\phi^2\right) \tag{1.168}$$

where we also replaced $\eta_{\mu\nu} \to g_{\mu\nu}$. Indeed, as we will discuss in the next section, these are also the equations that we obtain from a straightforward conversion of the scalar field Lagrangian density (1.164) to curved space-time.

However, we could also have chosen to write instead

$$\nabla^\mu \nabla_\mu \phi - m^2\phi - \xi R\phi = 0\,, \tag{1.169}$$

where ξ is a constant[14]. This is entirely consistent with the two principles given in the last section, namely that only the metric and quantities derived from it may appear, and that the equation should reduce to the one of SR in the appropriate limit. The choice $\xi = 0$ is called minimally coupled, for reasons that will become clear in the next section.

The electromagnetic field in special and general relativity

In Maxwell's theory of electromagnetism, the electric field $\mathbf{E}$ and the magnetic field $\mathbf{B}$ (spatial vectors) are combined into an antisymmetric space-time field tensor $F_{\mu\nu}$. For an observer moving with 4-velocity v^μ,

$$E_\mu = F_{\mu\nu}v^\nu \tag{1.170}$$

is interpreted as the electric field measured by that observer, while

$$B_\mu = -\frac{1}{2}\epsilon_{\mu\nu}{}^{\alpha\beta}F_{\alpha\beta}v^\nu \tag{1.171}$$

is the magnetic field. Here $\epsilon_{\mu\nu\alpha\beta}$ is the totally antisymmetric tensor, with norm $\epsilon_{\mu\nu\alpha\beta}\epsilon^{\mu\nu\alpha\beta} = -24$. Maxwell's equations then are simply

$$\partial^\mu F_{\mu\nu} = -4\pi j_\nu\,, \tag{1.172}$$
$$\partial_{[\mu}F_{\nu\lambda]} = 0\,, \tag{1.173}$$

where j_μ is the current density 4-vector. Due to the antisymmetry of $F_{\mu\nu}$ we have that

$$0 = \partial^\mu\partial^\nu F_{\mu\nu} = -4\pi\partial^\nu j_\nu\,, \tag{1.174}$$

[14] The value $\xi = 1/6$ is special since it makes the equation conformally invariant for $m = 0$, see e.g. appendix D of [1].

which corresponds to the conservation of charge, analogously to Eq. (1.121) implying the conservation of energy. In the presence of an electromagnetic (EM) field a charged particle experiences an external force (the Lorentz force), so that the equation of motion (1.111) becomes

$$u^\mu \partial_\mu u^\nu = \frac{q}{m} F^\nu{}_\lambda u^\lambda \,. \tag{1.175}$$

The electromagnetic stress-energy tensor is given by

$$T_{\mu\nu} = \frac{1}{4\pi}\left(F_{\mu\lambda}F_\nu{}^\lambda - \frac{1}{4}\eta_{\mu\nu}F_{\lambda\rho}F^{\lambda\rho}\right) . \tag{1.176}$$

If there is no external charge current then $\partial^\mu T_{\mu\nu} = 0$, otherwise the stress-energy tensor of the electromagnetic field is not independently conserved but the total stress-energy of the field and of the charged matter still is.

It is also possible to describe electromagnetism with the help of the *vector potential* A^μ,

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \,, \tag{1.177}$$

in which case Maxwell's equation are

$$\partial^\mu \left(\partial_\mu A_\nu - \partial_\nu A_\mu\right) = -4\pi j_\nu \,. \tag{1.178}$$

There is some gauge freedom in the vector potential, since adding a gradient of a scalar function, $\partial_\mu\chi$ to A_μ does not change $F_{\mu\nu}$ due to Eq. (1.177). We can use this freedom to impose the Lorenz gauge condition

$$\partial^\mu A_\mu = 0 \,, \tag{1.179}$$

which then allows to simplify (1.178) to

$$\partial^\mu \partial_\mu A_\nu = -4\pi j_\nu \,. \tag{1.180}$$

These equations could also be found from the Lagrangian density for the Maxwell's equation, which is given by the sum of the Lagrangian of the electromagnetic field $\mathcal{L}_{\rm EM}$ and, if charges are present, the interaction Lagrangian $\mathcal{L}_{\rm int}$,

$$\mathcal{L}_{\rm EM} = -\frac{1}{16\pi}F_{\mu\nu}F^{\mu\nu} = -\partial_{[\mu}A_{\nu]}\partial^{[\mu}A^{\nu]} \,, \quad \mathcal{L}_{\rm int} = A^\mu j_\mu \,, \tag{1.181}$$

where the pre-factor depends on the units, see e.g. [6]. The Euler-Lagrange equations in this case are

$$\frac{\partial\mathcal{L}}{\partial A_\mu} - \partial_\nu\left(\frac{\partial\mathcal{L}}{\partial(\partial_\nu A_\mu)}\right) = 0 \,, \tag{1.182}$$

which gives us the inhomogeneous Maxwell equation (1.172), the homogeneous one, Eq. (1.173), being automatically satisfied by A_μ.

The conversion of Maxwell's equations to GR is also straightforward,

$$\nabla^\mu F_{\mu\nu} = -4\pi j_\nu \,, \tag{1.183}$$
$$\nabla_{[\mu}F_{\nu\lambda]} = 0 \,, \tag{1.184}$$

and the stress-energy tensor (1.176) becomes

$$T_{\mu\nu} = \frac{1}{4\pi}\left(F_{\mu\lambda}F_\nu{}^\lambda - \frac{1}{4}g_{\mu\nu}F_{\lambda\rho}F^{\lambda\rho}\right) . \tag{1.185}$$

However, we could just as well have used the vector potential A_μ, that we can also introduce in GR (at least locally) thanks to Eq. (1.184). But the Maxwell equations in the Lorenz gauge, (1.180), contain two derivatives that in GR no longer commute. For this reason we find that we need to add a curvature term now when deriving the equation from (1.183) and using the (GR) Lorenz gauge condition $\nabla^\mu A_\mu = 0$ to remove the second term, leading to

$$\nabla^\mu \nabla_\mu A_\nu - R_\nu^\lambda A_\lambda = -4\pi j_\nu \,. \tag{1.186}$$

If we had applied the 'minimal substitution rule' directly to (1.180) we would not have seen the need to add the Ricci tensor. Here the reason why we prefer the equations given is that they imply charge conservation, $\nabla_\mu j^\mu = 0$[15]. But we see (as also in the case of the scalar field above) that the minimal substitution rule is not necessarily unique, and that in can be difficult to figure out which possible substitution is the 'correct' one, in the sense that it correctly describes the phenomena observed in nature.

We already wrote down the equation of motion for a charged test particle in GR in Eq. (1.124), we here just note that in terms of the trajectory $x^\mu(\tau)$ of the particle in a coordinate basis and proper time we could write it more explicitly as

$$\left(\frac{d^2x^\mu}{d\tau^2} + \Gamma^\mu_{\alpha\beta}\frac{dx^\alpha}{d\tau}\frac{dx^\beta}{d\tau}\right) = \frac{q}{m}F^\mu{}_\nu \frac{dx^\nu}{d\tau} \,. \tag{1.187}$$

Light propagation in SR and GR

An important example of an electromagnetic wave is light. To find a 'light-like' solution of Maxwell's equation, we look for wave-like oscillating solutions with a constant (or at most slowly varying) amplitude,

$$A_\mu = C_\mu e^{iS} \,, \tag{1.188}$$

where C_μ is here taken to be a constant vector field (i.e. a field with constant norm that is everywhere parallel to itself), and the function S is the *phase* of the wave, which we can take to be a real valued function. As light propagates in vacuum, we look for a solution of (1.180) with $j_\mu = 0$, which implies

$$\partial^\mu \partial_\mu S = 0 \,, \quad \partial_\mu S \partial^\mu S = 0 \,, \tag{1.189}$$

and imposing the Lorenz condition (1.179),

$$C_\mu \partial^\mu S = 0 \,. \tag{1.190}$$

The vector $\nabla^\mu f$ is normal (orthogonal) to the surfaces of constant f for any function f on any manifold with a metric, since for any vector $\mathbf{t}$ tangent to the surface of constant f we have $t^\mu \nabla_\mu f = 0$. Writing $k^\mu = \partial^\mu S$ for the vector normal to the surface of constant phase S (the wave vector in Fourier parlance), we see from (1.189) that k^μ is a null vector, $k^\mu k_\mu = 0$. Such a surface is called a *null hypersurface*, it has the property that the normal vector is tangent to the hypersurface. We also have that

$$0 = \partial_\nu(\partial_\mu S \partial^\mu S) = 2(\partial^\mu S)(\partial_\nu \partial_\mu S) = 2k^\mu \partial_\mu k_\nu \,, \tag{1.191}$$

i.e. the integral curves of k^μ are null geodesics. Actually this derivation also holds in curved space-times, replacing $\partial_\mu \to \nabla_\mu$, and it follows that all null hypersurfaces in a Lorentz space-time are generated by

[15] This can shown by proving that for any antisymmetric tensor field $F^{\mu\nu}$ we have that $\nabla_\mu \nabla_\nu F^{\mu\nu} = 0$ due to the symmetry of the Ricci tensor (or equivalently of the Christoffel symbols): $\nabla_\mu \nabla_\nu F^{\mu\nu} = [\nabla_\mu, \nabla_\nu]F^{\mu\nu} = 2R_{\mu\nu}F^{\mu\nu} = 0$.

null geodesics. The frequency of the wave (given by minus the rate of change of the phase) as measured by an observer with 4-velocity v^μ is

$$\omega = -v^\mu \partial_\mu S = -v^\mu k_\mu \,. \tag{1.192}$$

Since light rays are effectively the integral curves of the wave vector k^μ, we conclude that electromagnetic waves propagate on null geodesics as can be seen from Eq. (1.191). This can also be seen from the retarded Green's function which has only support on the past light cone. Equation (1.190) shows on the other hand that the polarisation vector, which is proportional to the amplitude C_μ, is always perpendicular to k^μ.

If the scale of variation of the electromagnetic field is much smaller than the scale of curvature, we should be able to write also in GR, as in SR, oscillating solutions of Maxwell's equation as

$$A_\mu = C_\mu e^{iS} \,. \tag{1.193}$$

Inserting this ansatz into Eq. (1.186) for $j_\nu = 0$ and neglecting both $\nabla_\mu \nabla^\mu C_\nu$ (as the amplitude is supposed to be a slowly varying function, contrary to the phase S) and the term with the Ricci tensor (as the curvature locally should be small) we obtain as before

$$\nabla_\mu S \nabla^\mu S = 0 \,. \tag{1.194}$$

As already mentioned before, also in curved space-time the surfaces of constant phase are null hypersurfaces, and $k_\mu = \nabla_\mu S$ is tangent to null geodesics. We can make this more explicit by using that $0 = \nabla_\mu (k^\nu k_\nu) = 2k^\nu \nabla_\mu k_\nu$. Since $k_\nu = \nabla_\nu S$ and according to (1.61) $\nabla_\mu \nabla_\nu f = \nabla_\nu \nabla_\mu f$, we have that

$$k^\nu \nabla_\nu k^\mu = 0 = \nabla_{\mathbf{k}} \mathbf{k} \,. \tag{1.195}$$

Thus, also in GR we expect that light travels on null geodesics, at least in this approximation, known as *geometrical optics approximation* or *eikonal approximation*.

If we define the polarisation vector as $f_\mu = C_\mu / C$ where $C = |\mathbf{C}|$ is the scalar amplitude, then the Lorenz condition, for our approximations, implies again

$$k_\mu C^\mu = k_\mu f^\mu = 0 \,, \tag{1.196}$$

i.e. the polarisation vector is perpendicular to the wave vector.

1.4.3 Minimally coupling matter to gravity

With the matter Lagrangians $\mathcal{L}_\phi$ and $\mathcal{L}_{\text{EM}}$ we can now extend the action principle in General Relativity to the non-vacuum case. For this, we add a suitable matter action S_m to the Einstein-Hilbert action. For example, the action to be used for the scalar field is translated from $\mathcal{L}_\phi$ given in (1.164) in the usual way, but we now also have to use the GR volume element,

$$S_\phi = \int_D \mathcal{L}_\phi d\mu = \int_D -\frac{1}{2} \left(\nabla_\mu \phi \nabla^\mu \phi + m^2 \phi^2 \right) \sqrt{-g}\, d^4x \,. \tag{1.197}$$

The same is true for the action of the electromagnetic field (1.181) where the translation is even simpler as we just need to use the right volume element (and of course the field tensor $F_{\mu\nu}$ of GR).

The matter action and the Einstein-Hilbert action can then be simply added, e.g. for the scalar field

$$S = S_{\text{EH}} + \alpha_\phi S_\phi \tag{1.198}$$

where α_ϕ is a (coupling) constant. This is called *minimal coupling* since we do not add any additional terms in the total action that explicitly involve gravity and matter, the only connection between gravity and matter is the use of the metric and of functions derived from it (like the covariant derivative and the volume element) in S_ϕ.

As $S_{\rm EH}$ and $\sqrt{-g}$ do not depend on ϕ, the equation of motion for ϕ (1.167) is unchanged, although we have to be a bit careful about the variation involving now covariant derivatives. Let us assume that the Lagrangian density is a function $\mathcal{L}(\psi, \nabla\psi, g)$ where ψ are the matter fields (tensors[16] like ϕ or A_μ). We thus have to solve

$$\delta \int_D \mathcal{L} d\mu \tag{1.199}$$

where the fields are varied inside of D so that the variation vanishes on the boundary ∂D. As usual, δ means differentiation with respect to a one-parameter family of field variations (see e.g. Appendix E of [1]). We then have

$$\delta \int_D \mathcal{L} d\mu = \int_D (\delta\mathcal{L}) d\mu = \int_D \left(\frac{\partial \mathcal{L}}{\partial \psi} \delta\psi + \frac{\partial \mathcal{L}}{\partial(\nabla\psi)} \delta(\nabla\psi) \right) d\mu \,. \tag{1.200}$$

As it is a derivative, δ commutes with partial differentiation in a coordinate basis, and also with the covariant derivative ∇ as can be seen from the coordinate expression for ∇, Eq. (1.58). We then have

$$\frac{\partial \mathcal{L}}{\partial(\nabla\psi)} \delta(\nabla\psi) = \nabla \left(\frac{\partial \mathcal{L}}{\partial(\nabla\psi)} \delta\psi \right) - \left(\nabla \frac{\partial \mathcal{L}}{\partial(\nabla\psi)} \right) \delta\psi \,. \tag{1.201}$$

The first term is a total divergence that vanishes on integration thanks to Gauss' theorem and the fact that we consider variations that vanish on the boundary, so that

$$\int_D \left(\frac{\partial \mathcal{L}}{\partial \psi} - \nabla \frac{\partial \mathcal{L}}{\partial(\nabla\psi)} \right) \delta\psi d\mu = 0 \,. \tag{1.202}$$

We thus find the Euler-Lagrange equations in GR, which are, not very surprisingly,

$$\frac{\partial \mathcal{L}}{\partial \psi_a} - \nabla_\mu \frac{\partial \mathcal{L}}{\partial(\nabla_\mu \psi_a)} = 0 \,, \tag{1.203}$$

where we explicitly wrote a generic index a for the matter fields.

We are also able to vary the total action with respect to $g^{\mu\nu}$. As the metric represents the dynamical degrees of freedom of gravity, we should obtain the non-vacuum Einstein equation, i.e. the variation of the matter action with respect to the metric should be proportional to the stress-energy tensor. The variation will be of the form

$$\delta \int_D \mathcal{L} d\mu = \int_D \left[\left(\frac{\partial \mathcal{L}}{\partial(\nabla_\lambda\psi)} \delta(\nabla_\lambda\psi) + \frac{\partial \mathcal{L}}{\partial g_{\mu\nu}} \delta g_{\mu\nu} \right) d\mu + \mathcal{L} \delta d\mu \right] \,. \tag{1.204}$$

For the variation of the measure $\delta d\mu$ we can use (1.161). We note that $\delta(\nabla_\lambda\psi)$ will in general be non-zero because of the Christoffel symbols that contain the metric. In Riemann normal coordinates at $p \in \mathcal{M}$ we can write, from (1.68),

$$\delta\Gamma^\mu_{\alpha\beta} = \frac{1}{2} g^{\mu\nu} \left((\delta g_{\nu\alpha})_{;\beta} + (\delta g_{\nu\beta})_{;\alpha} - (\delta g_{\alpha\beta})_{;\nu} \right) \,, \tag{1.205}$$

a tensorial equation that therefore will hold in any coordinate system and everywhere. After integration by parts we can then bring (1.204) into the form

$$\delta \int_D \mathcal{L} d\mu = \frac{1}{2} \int_D T^{\mu\nu} \delta g_{\mu\nu} d\mu \tag{1.206}$$

[16] Or spinors, but this is beyond the scope of this lecture, you can find more on that possibility in books like [1, 2].

where now only $g_{\mu\nu}$ is varied. We identify $T^{\mu\nu}$ in (1.206) with the stress-energy tensor, note that this $T^{\mu\nu}$ is automatically symmetric. Equivalently we can write

$$T_{\mu\nu} = -\frac{2}{\sqrt{-g}} \frac{\delta(\mathcal{L}\sqrt{-g})}{\delta g^{\mu\nu}} . \tag{1.207}$$

Note that these signs are consistent since from $g_{\mu\nu}g^{\nu\alpha} = \delta^\alpha_\mu$ we have that

$$\delta g_{\mu\nu}g^{\nu\alpha} + g_{\mu\nu}\delta g^{\nu\alpha} = 0 \quad \Leftrightarrow \quad T^{\mu\nu}\delta g_{\mu\nu} = -T_{\mu\nu}\delta g^{\mu\nu} . \tag{1.208}$$

The combined variation can then be found from

$$\delta \int_D \left(\frac{R}{16\pi G} + \mathcal{L} \right) d\mu = 0 \tag{1.209}$$

since with (1.160) and (1.206), and using also (1.208), this gives us the Einstein equation

$$G_{\mu\nu} = 8\pi G T_{\mu\nu} . \tag{1.210}$$

We see that (1.207) gives us a systematic way to compute the stress-energy tensor in GR once we have obtained the Lagrangian or action.

Electromagnetism

We can illustrate the variational procedure to obtain the stress energy for the electromagnetic field, by varying $\mathcal{L}_{\rm EM}$ from Eq. (1.181). We write it as

$$\mathcal{L}_{\rm EM} = -\frac{1}{16\pi} F_{\mu\nu}F_{\sigma\rho}g^{\mu\sigma}g^{\nu\rho} . \tag{1.211}$$

For the variation we then have, using again (1.161)

$$\delta \int_D \mathcal{L} d\mu = \int_D \left(\delta\mathcal{L} + \frac{1}{2}\mathcal{L}g^{\mu\nu}\delta g_{\mu\nu} \right) d\mu = \int_D \left(-\frac{1}{8\pi} F_{\mu\nu}F_{\sigma\rho}g^{\mu\sigma}\delta g^{\nu\rho} + \frac{1}{2}\mathcal{L}g^{\mu\nu}\delta g_{\mu\nu} \right) d\mu \tag{1.212}$$

From the left-hand relation of (1.208) we have that

$$\delta g^{\nu\rho} = -g^{\nu\beta}g^{\alpha\rho}\delta g_{\alpha\beta} , \tag{1.213}$$

and hence

$$\delta \int_D \mathcal{L} d\mu = \frac{1}{8\pi} \int_D \left(F_{\mu\nu}F_{\sigma\rho}g^{\mu\sigma}g^{\nu\beta}g^{\alpha\rho} - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}g^{\alpha\beta} \right) \delta g_{\alpha\beta} d\mu \tag{1.214}$$

Comparing this with (1.206) we find

$$T^{\alpha\beta} = \frac{1}{4\pi} \left(F^{\sigma\beta}F_\sigma{}^\alpha - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}g^{\alpha\beta} \right) \tag{1.215}$$

which is indeed the stress-energy tensor (1.185) if we lower the indices.

Varying $\mathcal{L}_{\rm EM}$, for $F_{\mu\nu} = A_{\nu;\mu} - A_{\mu;\nu} = A_{\nu,\mu} - A_{\mu,\nu}$, with respect to A_μ, is straightforward,

$$\delta \int_D \mathcal{L}_{\rm EM} d\mu = \frac{1}{8\pi} \int_D F^{\mu\nu}\delta(A_{\mu;\nu} - A_{\nu;\mu}) d\mu = \frac{1}{4\pi} \int_D F^{\mu\nu}\delta A_{\mu;\nu} d\mu = -\frac{1}{4\pi} \int_D F^{\mu\nu}_{;\nu}\delta A_\mu d\mu . \tag{1.216}$$

We see that, setting this equal to zero, we obtain the 'right' Maxwell equation for vacuum, Eq. (1.183) with $j_\nu = 0$.

To couple the electromagnetic field to charged matter or fields, we introduce an interaction Lagrangian $\mathcal{L}_{\rm int}$. The variation of the interaction Lagrangian is treated in an analogous way to the definition of the stress-energy tensor in Eq. (1.206). Here the variation defines the current density j^μ through

$$\delta \int_D \mathcal{L}_{\rm int} d\mu = \int_D j^\mu \delta A_\mu d\mu . \tag{1.217}$$

Maxwell's equation can then be found from the variation of the integral of $\mathcal{L}_{\rm EM} + \mathcal{L}_{\rm int}$ which has to vanish.

Non-minimal coupling

Finally we note that other prescriptions are possible. For example, we could have chosen to add a term containing both R and ϕ in the total action,

$$S = S_{\mathrm{EH}} + S_\phi - \int \frac{\xi}{2}\phi^2 R d\mu = \int \left[R\left(1 - \frac{\xi}{2}\phi^2\right) + \mathcal{L}_\phi \right] d\mu \,. \tag{1.218}$$

This is called *non-minimal coupling* since there is a direct coupling between ϕ and R. The equation of motion of the scalar field now becomes (1.169), which would have been acceptable from a special-relativistic point of view where R is zero, but from a Lagrangian point of view should be considered as a modification of GR, a so-called *scalar-tensor theory*. Another popular modification of the Einstein-Hilbert action consists in replacing R by a function of the Ricci scalar, $f(R)$. The construction of a modified Lagrangian density is often the starting point for construction of theories of gravity beyond GR. However, existence and stability of suitable solutions for describing the Universe, as well as compatibility with the very strong observational constraints on the propagation of gravitational waves and on the action of gravity in the solar system and double pulsar systems place stringent constraints on the space of possible modified actions.

1.5 How unique is General Relativity?

We have motivated the Einstein equations from an analogy with Newtonian gravity, and considerations based on the Equivalence Principle. But maybe we could have found a different theory that would also work? It turns out that, given fairly general conditions, and in 4 dimensions, this is actually not the case.

At the latest since the discovery of gravitational waves, we know that we are looking for a theory of spin-2 fields, which are described by a symmetric rank-2 tensor field $g_{\alpha\beta}$. We would expect the ten degrees of freedom $g_{\alpha\beta}$ to satisfy equations of the form

$$\mathcal{D}_{\mu\nu}[g] = T_{\mu\nu} \,, \tag{1.219}$$

where $\mathcal{D}_{\mu\nu}[g]$ is a tensor field that is constructed (point-wise) from $g_{\alpha\beta}$ and its first and second derivatives. The limitation to second derivatives is desirable as higher-order theories generically suffer from unacceptable instabilities and since for a higher order theory the initial conditions have to specify quantities beyond the initial "position" and "velocity" (value and first derivative of the functions).

We also want the tensor field to satisfy the identity

$$\nabla^\nu \mathcal{D}_{\mu\nu} = 0 \tag{1.220}$$

as we want the property $\nabla^\nu T_{\mu\nu} = 0$ to be a property of the field equations.

For such a theory, the *Lovelock theorem* holds [5]:

> A tensor $\mathcal{D}_{\mu\nu}[g]$ with the properties given above is in four dimensions a linear combination of the metric and the Einstein tensor,
>
> $$\mathcal{D}_{\mu\nu}[g] = aG_{\mu\nu} + bg_{\mu\nu} \,, \tag{1.221}$$
>
> where a and b are real constants.

In the following we will sketch the proof of the theorem under the additional condition that $\mathcal{D}_{\mu\nu}[g]$ should be linear in the second derivatives. This is however not necessary, in four dimensions this is a result. It is also not necessary to postulate the symmetry of $\mathcal{D}_{\mu\nu}$, this is again a result, and shows that the gravitational field of the type (1.219) cannot be coupled to a non-symmetric stress-energy tensor.

Following [2] we start by showing the following:
Let $K[g]$ be a functional which assigns to every smooth Lorentzian metric field g a tensor field K of arbitrary type such that the components of $K[g]$ at the point $p \in \mathcal{M}$ depend smoothly on $g_{\mu\nu}$ and its derivatives $g_{\mu\nu,\lambda}$, $g_{\mu\nu,\lambda\rho}$ at p. Then K is determined pointwise by g and the curvature tensor by elementary operations of tensor algebra.

Proof:
In Riemann normal coordinates (see Sect. 1.2.7) the geodesics are straight lines, $x^\mu(s) = sa^\mu$. Hence, since they depend only linearly on s, the geodesic equation is

$$\Gamma^\lambda_{\mu\nu}[sa^\rho]a^\mu a^\nu = 0\,, \tag{1.222}$$

for arbitrary constants a^μ. At $x = 0$, $\Gamma = 0$ and the derivatives of the metric vanish, $g_{\mu\nu,\lambda} = 0$. Furthermore we can bring the metric to normal form, $g_{\mu\nu}[0] = \epsilon_\mu \delta_{\mu\nu}$ with $\epsilon_\mu = \pm 1$. In a neighbourhood of $x = 0$ we can then write the metric as a series expansion,

$$g_{\mu\nu} = \epsilon_\mu \delta_{\mu\nu} + \frac{1}{2}\beta_{\mu\nu,\alpha\beta}x^\alpha x^\beta + \ldots\,. \tag{1.223}$$

Thus, for $\Gamma_{\lambda\mu\nu} = g_{\lambda\rho}\Gamma^\rho_{\mu\nu}$:

$$\Gamma_{\lambda\mu\nu} = \frac{1}{2}\left(g_{\lambda\mu,\nu} + g_{\lambda\nu,\mu} - g_{\mu\nu,\lambda}\right) = \frac{1}{2}\left(\beta_{\lambda\mu,\nu\sigma}x^\sigma + \beta_{\lambda\nu,\mu\sigma}x^\sigma - \beta_{\mu\nu,\lambda\sigma}x^\sigma\right) + \ldots\,. \tag{1.224}$$

As the geodesics are straight lines in Riemann normal coordinates, we have that $\Gamma^\lambda_{\mu\nu}[x]x^\mu x^\nu = 0$, and so also

$$\Gamma_{\lambda\mu\nu}[x]x^\mu x^\nu = 0\,. \tag{1.225}$$

The last two equations together imply

$$\left(\beta_{\lambda\mu,\nu\sigma} + \beta_{\lambda\nu,\mu\sigma} - \beta_{\mu\nu,\lambda\sigma}\right)x^\mu x^\nu x^\sigma = 0\,. \tag{1.226}$$

A not so short but relatively straightforward calculation that can be found e.g. on p. 76/77 of [2] shows that

$$R_{\mu\rho\nu\sigma} + R_{\nu\rho\mu\sigma} = -3\beta_{\mu\nu,\rho\sigma} = -3\left.\frac{\partial^2 g_{\mu\nu}}{\partial\rho\partial\sigma}\right|_{x=0}\,. \tag{1.227}$$

Hence, in normal coordinates we have

$$g_{\mu\nu} = \epsilon_\mu\delta_{\mu\nu} - \frac{1}{6}\left(R_{\mu\rho\nu\sigma} + R_{\nu\rho\mu\sigma}\right)x^\rho x^\sigma + \ldots = \epsilon_\mu\delta_{\mu\nu} - \frac{1}{3}R_{\mu\rho\nu\sigma}x^\rho x^\sigma + \ldots\,. \tag{1.228}$$

Thus, $R_{\mu\rho\nu\sigma}$, expressed in terms of Riemann normal coordinates, determines all the second derivatives of $g_{\mu\nu}$ at p. The first derivatives vanish. As $R_{\mu\rho\nu\sigma}$ is a tensor, the theorem follows.

We now assume that $\mathcal{D}_{\mu\nu}[g]$ is linear in the second derivatives (which is not actually a necessary assumption, but makes the proof much simpler). As $R^\mu_{\alpha\nu\beta}$ is linear in the second derivatives, the previous theorem implies that $\mathcal{D}_{\mu\nu}[g]$ must have the following form:

$$\mathcal{D}_{\mu\nu}[g] = c_1 R_{\mu\nu} + c_2 R g_{\mu\nu} + c_3 g_{\mu\nu} \tag{1.229}$$

with constants c_1, c_2 and c_3. As in the motivation of GR in Section 1.3.4 we require $\mathcal{D}^{\mu\nu}_{;\nu} = 0$, which implies because of the contracted Bianchi identity $G^{\mu\nu}_{;\nu} = 0$ that

$$\mathcal{D}^{\mu\nu}_{;\nu} = \left(c_2 + \frac{1}{2}c_1\right)\left(g^{\mu\nu}R\right)_{;\nu} = 0\,. \tag{1.230}$$

Therefore $c_2 + c_1/2 = 0$ and $\mathcal{D}_{\mu\nu}$ must be of the form (1.221), which implies that the field equations can be written as

$$G_{\mu\nu} = \kappa T_{\mu\nu} - \Lambda g_{\mu\nu} \,, \tag{1.231}$$

where Λ and κ are constants. κ can be fixed by considering the Newtonian limit, as in Section 1.3.4, which implies that $\kappa = 8\pi G$. The constant Λ is called the *cosmological constant*. It was originally introduced by Einstein when he applied GR for the first time to cosmology, and was looking for a static solution.

Writing the cosmological constant on the right hand side of the field equations, as in Eq. (1.231), we can consider $\Lambda g_{\mu\nu}$ as the stress-energy tensor of a perfect fluid with energy density and pressure

$$\rho_\Lambda = \frac{\Lambda}{8\pi G} \,, \quad P_\Lambda = -\rho_\Lambda \,. \tag{1.232}$$

Formally this corresponds to the properties of a vacuum energy.

Today we know that the Universe is not static but is expanding. However, cosmological observations indicate that Λ is not actually zero, but very small relative to naive contributions to a vacuum energy from particle physics, about $4 \times 10^{-66}\mathrm{eV}^2$ in natural units. One possibility is that the contributions from quantum field theory cancel nearly exactly with the bare cosmological constant allowed by Lovelock's theorem. The reason for such a cancellation is not understood at all, this is known as the *cosmological constant problem*. For the purpose of this lecture, since Λ is observationally constrained to be so small, we can safely neglect it for the solutions that we will consider. This is however not the case in cosmology, where the cosmological constant is actually the dominant contribution to the total energy density today.

The Lagrangian density $\mathcal{L}[g]$, whose Euler-Lagrange derivative is equal to (1.231), is given by

$$\mathcal{L}[g] = R[g] - 2\Lambda \,. \tag{1.233}$$

Thus Lovelock's theorem implies that in four dimensions, a linear function of the Riemann scalar $R[g]$ is, up to a total divergence, the most general Lagrangian density whose Euler-Lagrange derivative contains no higher than second order derivatives of the metric field.

To get around Lovelock's theorem and to construct metric theories of gravity that are different from GR we have to break one or more of the assumptions of the theorem, like

- Add other fields in addition to the metric tensor (e.g. a scalar field with non-minimal coupling).
- Consider manifolds with more than four space-time dimensions (string theory, 'braneworlds')
- Allow for higher than second derivatives of the metric in the field equations.
- Consider field equations that are not given by a rank 2 tensor field.
- Give up locality or diffeomorphism invariance.

Chapter 2

Weak gravitational fields and gravitational waves

Most gravitational fields in the universe are weak, they become strong only near compact astrophysical objects of sizes of a about a solar mass to a few billion solar masses (neutron stars at the low mass end, then black holes), as shown in Fig. 2.1. For example, the gravitational field in the solar system is of the order

$$\phi \lesssim \frac{GM_\odot}{R_\odot c^2} \approx 10^{-6} \,. \tag{2.1}$$

To describe that part of the parameter space we can use perturbation theory around a suitable 'background metric', which is the topic of this chapter (except that we will limit ourselves to a flat background metric, a different choice will be used in cosmology). We will mostly follow the presentation of [2].

An important technical aspect that we will encounter in this section concerns the dependence of results on the choice of coordinate system or 'gauge'. As mentioned in the first chapter, the symmetry group of GR is the group of diffeomorphisms of $\mathbb{R}^4$. However, as we will explicitly find out, invariance of the physical results under coordinate transformations does not mean that the equations will *look* the same, only that their actual physical content has to be the same. It can be quite difficult to understand e.g. how many degrees of freedom a gravitational wave has.

2.1 Linearized theory of gravity

We assume that the metric describing space-time is nearly flat, at least in the region of interest. Then there exists a coordinate system so that

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu} \,, \quad |h_{\mu\nu}| \ll 1 \,. \tag{2.2}$$

As mentioned, in the solar system $|h_{\mu\nu}|$ is of the order of 10^{-6}. We will however allow for the field to vary rapidly, i.e. we do not assume that we can neglect time derivatives. This means that we are not necessarily in a quasi-Newtonian setting.

We then expand the field equations in powers of $h_{\mu\nu}$ and keep only the linear terms. In the Ricci tensor (1.153) the terms involving squares of the Christoffel symbols disappear to this order as they are either zero or quadratic in $h_{\mu\nu}$, so that we only keep

$$R_{\mu\nu} = \partial_\rho \Gamma^\rho_{\mu\nu} - \partial_\nu \Gamma^\rho_{\mu\rho} \,. \tag{2.3}$$

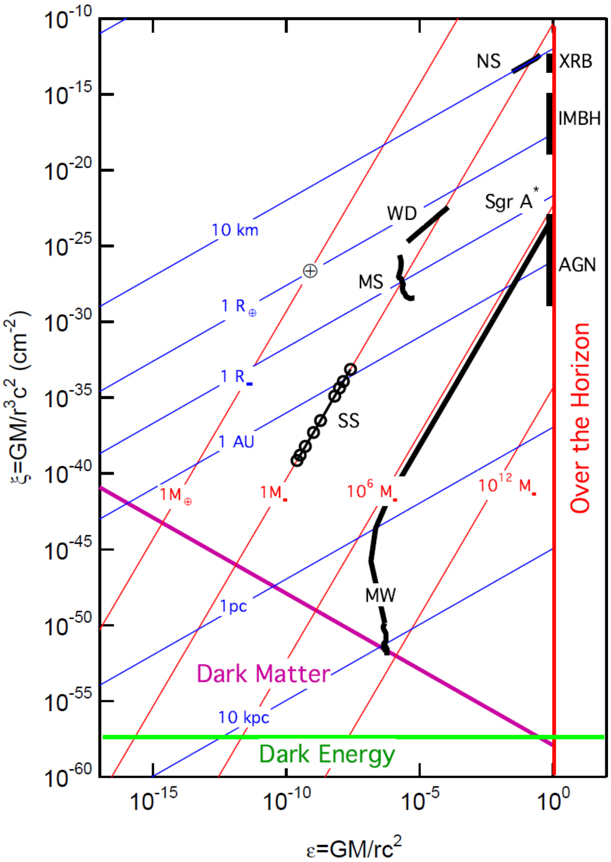

Figure 2.1: A graph that shows the strength of the gravitational field for various astrophysical and cosmological objects. The x-axis measures the gravitational potential $\phi = GM/r$ (called ϵ in the figure) which becomes unity at a black hole horizon, while the y-axis shows the curvature (in 1/cm^2). Most objects are situated at very small values of the potential, only compact objects like black holes and neutron stars, and some extreme astrophysical objects likely associated with supermassive black holes like Active Galactic Nuclei (AGN), are located close to the vertical line at $\phi = 1$. Figure from D. Psaltis, Living Rev. Relativ. (2008) 11: 9. https://doi.org/10.12942/lrr-2008-9.

For the Christoffel symbols (1.68) we can use the linearized approximation where we exploit again that $\partial_\lambda \eta_{\mu\nu} = 0$,

$$\Gamma^\alpha_{\mu\nu} = \frac{1}{2}\eta^{\alpha\beta}\left(h_{\mu\beta,\nu} + h_{\beta\nu,\mu} - h_{\mu\nu,\beta}\right) . \tag{2.4}$$

In this chapter we will use the convention that indices are raised and lowered with $\eta_{\mu\nu}$, not the full metric $g_{\mu\nu}$, with the exception of indices of the metric itself (to ensure that $g^\nu_\mu = \delta^\nu_\mu$, at the level of our approximation this implies that the inverse metric is given by $g^{\mu\nu} = \eta^{\mu\nu} - h^{\mu\nu}$). The linearized Christoffel symbols can then be written as

$$\Gamma^\alpha_{\mu\nu} = \frac{1}{2}\left(h^\alpha_{\mu,\nu} + h^\alpha_{\nu,\mu} - h_{\mu\nu}{}^{,\alpha}\right) , \tag{2.5}$$

and the Ricci tensor becomes

$$R_{\mu\nu} = \frac{1}{2}\left(h^\lambda_{\mu,\lambda\nu} - \Box h_{\mu\nu} - h^\lambda_{\lambda,\mu\nu} + h^\lambda_{\nu,\lambda\mu}\right) , \tag{2.6}$$

for $\Box = \partial_\lambda \partial^\lambda$. The Ricci scalar is then

$$R = h^{\lambda\sigma}{}_{,\lambda\sigma} - \Box h , \tag{2.7}$$

where $h \equiv h^\lambda_\lambda = \eta^{\alpha\beta} h_{\alpha\beta}$. The linearized Einstein tensor is now

$$2G_{\mu\nu} = -\Box h_{\mu\nu} - h_{,\mu\nu} + h^\lambda_{\mu,\lambda\nu} + h^\lambda_{\nu,\lambda\mu} + \eta_{\mu\nu}\Box h - \eta_{\mu\nu} h^{\lambda\sigma}{}_{,\lambda\sigma}\,, \tag{2.8}$$

and the linearized field equations

$$\Box h_{\mu\nu} + h_{,\mu\nu} - h^\lambda_{\mu,\lambda\nu} - h^\lambda_{\nu,\lambda\mu} - \eta_{\mu\nu}\Box h + \eta_{\mu\nu} h^{\lambda\sigma}{}_{,\lambda\sigma} = -16\pi G T_{\mu\nu}\,. \tag{2.9}$$

We also note that for any quantity that is already first order in $h_{\mu\nu}$ the covariant derivative is simply the normal derivative as the Christoffel symbols are also first order in the perturbation. The contracted Bianchi identity is still verified at the level of our approximation,

$$G^{\mu\nu}{}_{,\nu} = 0\,. \tag{2.10}$$

The contracted Bianchi identity and the linearized field equation imply that

$$T^{\mu\nu}{}_{,\nu} = 0\,. \tag{2.11}$$

This means that in this approximation the field generated by $T_{\mu\nu}$ does not affect the source. For example, for incoherent dust, $T^{\mu\nu} = \rho u^\mu u^\nu$, for which Eq. (2.11) implies $u^\nu u^\mu{}_{,\nu} = 0$ (in addition to $(\rho u^\mu)_{,\mu} = 0$, cf. discussion after Eq. (1.115)). This is the equation for a straight geodesic, $du^\mu/ds = 0$, in other words integral curves of u^μ are straight lines, which is obviously not a good description of nature. When we look at the Newtonian limit in the next section we will thus have to partially go beyond the linear approximation, for example with an iterative procedure whereby we determine $h_{\mu\nu}$ from (2.9), using the 'flat' $T_{\mu\nu}$ based on $\eta_{\mu\nu}$, and then use $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$ to compute the evolution of physical systems (particles, fields) from the geodesic equation and the field equations of motion.

We can simplify some of the expressions by using

$$\gamma_{\mu\nu} \equiv h_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu} h \tag{2.12}$$

instead of $h_{\mu\nu}$. The inverse relation is

$$h_{\mu\nu} = \gamma_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}\gamma\,, \tag{2.13}$$

where we set $\gamma \equiv \gamma^\lambda_\lambda$. In this variable the linearized field equations (2.9) become

$$-\Box\gamma_{\mu\nu} - \eta_{\mu\nu}\gamma_{\alpha\beta}{}^{,\alpha\beta} + \gamma_{\mu\alpha,\nu}{}^{,\alpha} + \gamma_{\nu\alpha,\mu}{}^{,\alpha} = 16\pi G T_{\mu\nu}\,. \tag{2.14}$$

Since we have the flat 'background' metric $\eta_{\mu\nu}$ we can think of the linearized theory as a Lorentz-covariant field theory in flat space for the fields $h_{\mu\nu}$. We can then look at the transformation of $h_{\mu\nu}$ under a global Lorentz transformation,

$$x^\mu = \Lambda^\mu_\nu x'^\nu\,, \quad \Lambda^T\eta\Lambda = \eta\,, \tag{2.15}$$

where $\eta = \mathrm{diag}(-1,1,1,1)$. We can write, using the transformation properties of the metric tensor $g_{\mu\nu}$ under coordinate transformations,

$$\eta_{\alpha\beta} + h'_{\alpha\beta} = \frac{\partial x^\mu}{\partial x'^\alpha}\frac{\partial x^\nu}{\partial x'^\beta} g_{\mu\nu} = \Lambda^\mu_\alpha\Lambda^\nu_\beta(\eta_{\mu\nu} + h_{\mu\nu}) = \eta_{\alpha\beta} + \Lambda^\mu_\alpha\Lambda^\nu_\beta h_{\mu\nu}\,. \tag{2.16}$$

We find that $h_{\mu\nu}$ transforms as a tensor under the Lorentz group,

$$h'_{\alpha\beta} = \Lambda^\mu_\alpha\Lambda^\nu_\beta h_{\mu\nu}\,. \tag{2.17}$$

It is actually possible to start from this linear spin-2 field theory in flat Minkowski space, and to implement the iterative procedure mentioned above to derive a non-linear completion. This non-linear completion then leads to field equations that turn out to be equivalent to Einstein's equations. GR could therefore have been constructed as a field theory on Minkowski space, however even in this case it is natural to combine the flat background metric and the field to a dynamical metric and to reinterpret the theory geometrically; see Section 2.2.4 of [2] for more details and references.

There is also a gauge group for the linearized theory (as in many field theories, like electrodynamics), in the sense that the linearized Einstein tensor (2.8) is invariant under gauge transformations of the form

$$h_{\mu\nu} \to h'_{\mu\nu} = h_{\mu\nu} + \xi_{\mu,\nu} + \xi_{\nu,\mu}\,, \tag{2.18}$$

where ξ^μ is a vector field, as can be verified with a straightforward direct calculation. This transformation can be considered as an *infinitesimal coordinate transformation*

$$x'^\mu = x^\mu - \xi^\mu(x)\,, \tag{2.19}$$

where ξ^μ is an infinitesimal vector field (infinitesimal to preserve the condition (2.2) without which the linearized theory makes no sense), the minus sign is arbitrary and chosen so that the result agrees with (2.18). For most quantities the infinitesimal change can be neglected to the order of our approximation, while for the metric we find, based on the tensor transformation law[1]

$$g_{\mu\nu}(x) = \frac{\partial x'^\sigma}{\partial x^\mu}\frac{\partial x'^\rho}{\partial x^\nu} g'_{\sigma\rho}(x')\,, \quad \frac{\partial x'^\alpha}{\partial x^\beta} = \delta^\alpha_\beta - \xi^\alpha{}_{,\beta}\,, \tag{2.20}$$

so that

$$g_{\mu\nu}(x) = g'_{\mu\nu}(x') - g'_{\sigma\nu}(x')\xi^\sigma{}_{,\mu} - g'_{\mu\sigma}(x')\xi^\sigma{}_{,\nu}\,. \tag{2.21}$$

To first order in ξ^μ and $h_{\mu\nu}$ we find

$$h_{\mu\nu}(x) = h'_{\mu\nu}(x) - \xi_{\mu,\nu} - \xi_{\nu,\mu}\,. \tag{2.22}$$

This gauge freedom can be used to simplify the field equations: We can always find a gauge, called *Hilbert gauge*, such that

$$\gamma^{\alpha\beta}{}_{,\beta} = 0\,. \tag{2.23}$$

To see this, we note that (2.18) corresponds to

$$\gamma'_{\mu\nu} = \gamma_{\mu\nu} + \xi_{\mu,\nu} + \xi_{\nu,\mu} - \eta_{\mu\nu}\xi^\lambda{}_{,\lambda}\,, \tag{2.24}$$

which implies

$$\gamma'^{\mu\nu}{}_{,\nu} = \gamma^{\mu\nu}{}_{,\nu} + \Box\xi^\mu + \xi^{\nu,\mu}{}_{,\nu} - \xi^\lambda{}_{,\lambda}{}^{,\mu} = \gamma^{\mu\nu}{}_{,\nu} + \Box\xi^\mu\,. \tag{2.25}$$

We can always find a solution ξ^μ of the equation

$$\Box\xi^\mu = -\gamma^{\mu\nu}{}_{,\nu} \tag{2.26}$$

which then gives the γ that satisfies the Hilbert gauge condition. In Hilbert gauge, the field equations (2.14) become simply

$$\Box\gamma_{\mu\nu} = -16\pi G T_{\mu\nu}\,. \tag{2.27}$$

[1] We could also have framed the discussion of gauge invariance more geometrically in terms of the diffeomorphisms generated by a vector field as briefly discussed on page 13. The discussion can then be generalized to perturbations around arbitrary, curved background metrics $\bar{g}$. This is very important for perturbation theory in cosmology, we will however not go into more detail here.

This looks like a sourced wave equation for all components $h_{\mu\nu}$, which would be puzzling as we expect to have only *two* propagating degrees of freedom for a massless spin-2 field. This is actually an artefact of our gauge choice. A more careful investigation, that we will perform in Section 2.3, will find that there are, as expected, six degrees of freedom in $h_{\mu\nu}$, and that four of them satisfy Poisson-type equations, leaving two with wave-like equations. Those two degrees of freedom represent gravitational waves. This illustrates that care must be taken when interpreting results in specific gauges.

We can construct a formal solution of Eq. (2.27) that satisfies also (2.23) with the Green's function method (see e.g. the chapter on Green's functions and exercise 12 of the complex integration chapter in our mathematical methods course). The most general solution is

$$\gamma_{\mu\nu} = 16\pi G D_R * T_{\mu\nu} + s_{\mu\nu}\,, \tag{2.28}$$

where $s_{\mu\nu}$ is a solution of the homogeneous wave equation, and D_R is the retarded Green's function

$$D_R(x) = \frac{1}{4\pi|\mathbf{x}|}\delta(x^0 - |\mathbf{x}|)\,. \tag{2.29}$$

We can explicitly write the retarded solution as

$$\gamma_{\mu\nu}(x) = 4G\int \frac{\delta(x^0 - y^0 - |\mathbf{x}-\mathbf{y}|)}{|\mathbf{x}-\mathbf{y}|}T_{\mu\nu}(y)d^4y = 4G\int \frac{T_{\mu\nu}(x^0 - |\mathbf{x}-\mathbf{y}|, \mathbf{y})}{|\mathbf{x}-\mathbf{y}|}d^3y\,, \tag{2.30}$$

which we interpret as the field generated by the source $T_{\mu\nu}$, while the homogeneous solution represents gravitational waves that are present in the initial conditions, which we can think of as waves coming from infinity. From the Green's function solution above we conclude that gravitational effects propagate at the speed of light, as in electrodynamics.

To show that the solution (2.30) satisfies the Hilbert gauge condition, we use the variable transformation $y \to x - y$, which moves the x-dependence from the Green's function to the stress-energy tensor (effectively we use the symmetry under exchange of the functions of a convolution). Then we have explicitly that

$$\partial_\nu(D_R * T_{\mu\nu}) = D_R * \partial_\nu T_{\mu\nu} = 0\,. \tag{2.31}$$

The solution to the homogeneous equation will be discussed later, in Section 2.3.

2.2 Nearly-Newtonian fields

We call a nearly-Newtonian source one where velocities and stresses are small, $T_{00} \gg |T_{0j}|, |T_{ij}|$, and where we can neglect retardation effects. Then from (2.30) we obtain

$$\gamma_{00} = -4\phi\,, \quad \gamma_{0j} = \gamma_{ij} = 0 \tag{2.32}$$

where ϕ is the Newtonian potential

$$\phi(\mathbf{x}) = -G\int \frac{T_{00}(t,\mathbf{y})}{|\mathbf{x}-\mathbf{y}|}d^3y\,. \tag{2.33}$$

Translating this back to the metric $g_{\mu\nu}$, using (2.13), we find

$$g_{00} = -(1+2\phi)\,, \quad g_{0i} = 0\,, \quad g_{ij} = (1-2\phi)\delta_{ij}\,, \tag{2.34}$$

or

$$ds^2 = -(1+2\phi)dt^2 + (1-2\phi)d\mathbf{x}^2\,. \tag{2.35}$$

If the source is localized in space with total mass M then at large distances the monopole contribution will dominate and the metric will approach

$$ds^2 = -\left(1 - 2\frac{GM}{r}\right) dt^2 + \left(1 + 2\frac{GM}{r}\right) d\mathbf{x}^2 \,. \tag{2.36}$$

The neglected terms are of the order ϕ^2, $v\phi$ and $T_{ij}/T_{00}\phi$, in the solar system those are of order 10^{-12}, vs $\phi \sim 10^{-6}$. We see that the spatial part of the nearly-Newtonian metric is actually non-Euclidean.

Newtonian equations of motion

The motion of test particles is governed by the geodesic equation for the world line $x^\mu(t)$ in global inertial coordinates,

$$\frac{d^2x^\mu}{d\tau^2} + \Gamma^\mu_{\rho\sigma}\frac{dx^\rho}{d\tau}\frac{dx^\sigma}{d\tau} = 0\,. \tag{2.37}$$

For non-relativistic velocities we can approximate $dx^\mu/dt = (1,0,0,0)$, and $\tau \approx x^0 = t$. This leads to

$$\ddot{x}^\mu = -\Gamma^\mu_{00} \tag{2.38}$$

and from (2.5) we have that

$$\Gamma^i_{00} = -\frac{1}{2}\partial^i h_{00} = \partial^i\phi\,. \tag{2.39}$$

We see that the motion of non-relativistic test-bodies in a nearly Newtonian setting is governed by $\ddot{\mathbf{x}} = -\nabla\phi$, which is indeed the Newtonian equation of motion. We emphasize however once more that the interpretation of gravity in Newtonian theory and in GR is very different. In Newtonian theory massive bodies create a gravitational field that attracts other massive objects. The Earth is then going round the sun because of the gravitational field of the sun that creates a force and accelerates the Earth. If the Earth had no 'gravitational charge' it would follow a straight line, which is the trajectory of inertial observers in Newtonian theory. In GR however, also in the weak field limit discussed here, the mass of the sun curves space-time. The Earth, and any inertial observers, follow the curved geodesics of space-time, they are not accelerated. The equivalence principle is built into the theory, distinguishing between 'inertial mass' and 'gravitational mass' makes no sense. Instead, any observers who want to follow the straight geodesics of the $\eta_{\mu\nu}$ 'background' metric have to accelerate.

Propagation of light rays

We briefly revisit the geometrical optics approximation for light rays in an almost Newtonian metric. In this case we can write the eikonal equation (1.194), $g^{\mu\nu}\nabla_\mu S\nabla_\nu S = 0$, as

$$-(1-2\phi)(\partial_t S)^2 + (1+2\phi)(\nabla S)^2 = 0\,. \tag{2.40}$$

From (1.192) we have that, for small velocities, $\omega = -\partial_t S$, and since ϕ is time independent we write

$$S(\mathbf{x},t) = u(\mathbf{x}) - \omega t \tag{2.41}$$

and obtain (up to higher orders in ϕ)

$$(\nabla u)^2 = n^2\omega^2\,, \quad n = 1 - 2\phi\,. \tag{2.42}$$

This is a typical form of the eikonal equation in ray optics with refraction index n (see below). The conclusion is that light rays follow a trajectory as if they propagated in an inhomogeneous medium with a spatially varying index of refraction given by $1 - 2\phi$. This can be used to describe gravitational lensing, although we will follow a more direct approach in Section 2.5.

As the reader is maybe not familiar with the eikonal equation, we provide a brief derivation to place it into context. We consider for simplicity a scalar wave equation with velocity $v(\mathbf{x}) = c/n(\mathbf{x})$, corresponding to a light wave in a medium with position-dependent refractive index $n(\mathbf{x})$. The wave follows a wave equation

$$\frac{\partial^2 \phi}{\partial t^2} - \frac{c^2}{n^2}\nabla^2\phi = 0\,. \tag{2.43}$$

As above, we assume that the wave is nearly a plane wave (so that it has a clear ray-optics interpretation),

$$\phi(\mathbf{x}, t) = Ae^{iS}\,, \quad S(\mathbf{x}, t) = u(\mathbf{x}) - \omega t\,. \tag{2.44}$$

Inserting this ansatz into the wave equation (2.43), and neglecting again derivatives of A as we assume that amplitude varies much more slowly than the phase S, we find

$$\phi\left[-\omega^2 + \frac{c^2}{n^2}(\nabla u(\mathbf{x}))^2 - i\frac{c^2}{n^2}\nabla^2 u(\mathbf{x})\right] = 0\,. \tag{2.45}$$

The real and the imaginary part of this expression in square brackets vanish separately, and the real part (setting $c = 1$) gives us the (scalar) eikonal equation (2.42). This equation thus does indeed describe ray optics in a medium with (slowly varying) refractive index $n(\mathbf{x})$.

Gravitomagnetic field and Lense-Thirring precession

Let us look what happens if further terms of $T_{\mu\nu}$ become relevant, to see deviations from Newtonian dynamics. We assume that the stresses T_{ij} can be neglected, but we keep T_{0j} (which in general is v/c suppressed relative to T_{00}). The field equations (2.27) now are

$$\Box\gamma_{ij} = 0\,, \quad \Box\gamma_{0\mu} = -16\pi G T_{0\mu}\,. \tag{2.46}$$

We introduce a 'gravitational vector potential' $A_\mu \equiv \frac{1}{4}\gamma_{0\mu}$ that then satisfies a Maxwell-like equation (compare to Eq. (1.180)),

$$\Box A_\mu = -4\pi j_\mu\,, \tag{2.47}$$

where $j_\mu = GT_{0\mu}$ is proportional to the mass-energy current density. It is natural to define 'gravitational electric and magnetic fields' $\mathbf{E}$ and $\mathbf{B}$ as in electrodynamics.

Let us assume that the time derivatives of $\gamma_{\mu\nu}$ can be neglected, a quasi-stationary approximation that appears natural for sources with $v \ll c$. Then $\Delta\gamma_{ij} = 0$ everywhere in space, which implies $\gamma_{ij} = 0$, and hence A_μ describes the gravitational field completely. It is given by

$$A_0 = -\phi\,, \quad A_i(\mathbf{x}) = G\int \frac{T_{0i}(\mathbf{y})}{|\mathbf{x} - \mathbf{y}|}d^3y\,. \tag{2.48}$$

Going back to the metric $g_{\mu\nu}$ we find

$$g_{00} = -1 + 2A_0\,, \quad g_{0i} = 4A_i\,, \quad g_{ij} = (1 + 2A_0)\delta_{ij}\,. \tag{2.49}$$

The geodesic equation of motion for a (test) particle can be found from the variational principle

$$\delta\int\sqrt{-g_{\mu\nu}\frac{dx^\mu}{dt}\frac{dx^\nu}{dt}}dt = 0\,. \tag{2.50}$$

Writing $dx^\mu/dt = (1, \mathbf{v})$ and neglecting the higher-order term $A_0 v^2$, we find

$$g_{\mu\nu}\frac{dx^\mu}{dt}\frac{dx^\nu}{dt} = -1 + \mathbf{v}^2 + 2A_0 + 8\mathbf{A}\cdot\mathbf{v}\,. \tag{2.51}$$

This is actually closely analogous to electrodynamics, where the relativistic Lagrangian for a charged particle in an external electromagnetic field is[2]

$$\mathcal{L} = -m\sqrt{1-\mathbf{v}^2} - e\phi + e\mathbf{A}\cdot\mathbf{v}\,. \tag{2.52}$$

We see that, to leading order and up to the numerical pre-factor[3] 4 of the term involving $\mathbf{A}$ this is the same for $e = m = 1$, and the Euler-Lagrange equations give thus a result corresponding to (1.175), again in our approximation,

$$\ddot{\mathbf{x}} = \mathbf{E} + 4\dot{\mathbf{x}}\wedge\mathbf{B}\,, \tag{2.53}$$

except that now $\mathbf{E}$ and $\mathbf{B}$ represent the *gravitoelectric* and *gravitomagnetic* fields that act on neutral particles. The left-hand side of this equation of motion is the geodesic equation for the flat 'background' space-time metric $\eta_{\mu\nu}$ and the right-hand side are the gravitational effects induced by $T_{\mu\nu}$. The gravitomagnetic part corresponds to frame-dragging, an effect that is not present in Newtonian gravity. Efforts to measure it have been ongoing since decades, most recently with the Gravity Probe B space mission, but there is some controversy whether this has been done conclusively due to the small size of the frame-dragging in an orbit around Earth. The situation is however very different close to rotating black holes, a topic briefly touched upon toward the end of the course.

In the same way we can find the equivalent of the spin precession formula $\dot{\mathbf{S}} = \boldsymbol{\mu}\wedge\mathbf{B}$, where in electrodynamics $\boldsymbol{\mu} = e\mathbf{S}/(2m)$, by simply using the replacement $e \to m$ and $\mathbf{B} \to 4\mathbf{B}$. The resulting formula is called *Lense-Thirring precession* of a gyroscope. It shows that due to the gravitomagnetic field a rotating object, like a star, drags along a local inertial system, with precession frequency

$$\boldsymbol{\Omega}_{\mathrm{LT}} = -2\nabla\wedge\mathbf{A}\,. \tag{2.54}$$

2.3 Gravitational waves in the linearized theory

We now consider the linearized theory in vacuum. In the Hilbert gauge the field equations are then (2.27),

$$\Box\gamma_{\mu\nu} = 0\,. \tag{2.55}$$

As already mentioned after Eq. (2.27), this looks a like a wave equation for all components of $\gamma_{\mu\nu}$. Understanding how many propagating degrees we really have, and finding a gauge which makes this explicit, is the main technical difficulty in this section.

Actually, the Hilbert gauge does not fix the gauge completely. As we will now show, for the free field it is possible to find a gauge so that the trace vanishes additionally,

$$\gamma = 0\,. \tag{2.56}$$

[2]See e.g. Chapter 12 of [6]: The term $-m\sqrt{1-v^2}$ is the free particle Lagrangian, while the interaction Lagrangian is, as before, $A^\mu j_\mu$, but now A^μ is an external field and $j_\mu = e\,dx^\mu/dt$ is the current due to the moving charged particle.

[3]Looking at the equations, the factor arises on the one hand from the translation $\gamma_{\mu\nu} \to h_{\mu\nu}$ as the 'gravitational Maxwell equation' (2.47) is in terms of $\gamma_{\mu\nu}$, and on the other hand from the factor of two in the $0-i$ component of (2.51) as the metric is a symmetric tensor. At least the latter factor of 2 is directly a consequence of the tensorial nature of gravity, while electrodynamics is of a fundamentally vectorial nature. This also implies another difference: the vector potential A_μ and the current j_μ of electrodynamics transform as 4-vectors under Lorentz transformations, while the gravity equivalent do not: the mass-energy current density j_μ is not a 4-vector, but instead the part $T_{0\mu}$ of a tensor, as is $\gamma_{0\mu}$. There is additionally a sign-difference hidden in the equations, as gravity is attractive while in electrodynamics equal charges repel.

Under a change of gauge, $\gamma_{\mu\nu}$ changes according to (2.24),

$$\gamma_{\mu\nu} \to \gamma_{\mu\nu} + \xi_{\mu,\nu} + \xi_{\nu,\mu} - \eta_{\mu\nu}\xi^{\lambda}{}_{,\lambda}\,, \tag{2.57}$$

which implies

$$\gamma \to \gamma - 2\xi^{\lambda}{}_{,\lambda}\,, \quad \gamma^{\mu\nu}{}_{,\nu} \to \gamma^{\mu\nu}{}_{,\nu} + \Box\xi^{\mu}\,. \tag{2.58}$$

To preserve the Hilbert condition (2.23) we therefore have to impose

$$\Box\xi^{\mu} = 0\,. \tag{2.59}$$

So if $\gamma \neq 0$ we have to find a ξ^{μ} that satisfies this constraint as well as $2\xi^{\lambda}{}_{,\lambda} = \gamma$. The constraint (2.59) requires for consistency that $\Box\gamma = 0$ (as $\Box\gamma$ is invariant under the remaining gauge freedom) which is generally only true in vacuum. But if $\Box\gamma = 0$ then we can always find a ξ^{μ} that fulfils the conditions given: If ϕ is a scalar field such that $\Box\phi = 0$ then there exists a vector field ξ^{μ} with the properties $\Box\xi^{\mu} = 0$ and $\xi^{\mu}{}_{,\mu} = \phi$. We will now construct this vector field explicitly.

Let η^{μ} be a solution of $\eta^{\mu}{}_{,\mu} = \phi$. Such a solution exists, since one can use $\eta_{\mu} = U_{,\mu}$ for any function U that solves $\Box U = \phi$. Now choose $\zeta^{\mu} = \Box\eta^{\mu}$. From a generalisation of the Helmholtz decomposition we have that for a vector field with $\zeta^{\mu}{}_{,\mu} = 0$ there exists an antisymmetric tensor field $f^{\mu\nu}$ such that $\zeta^{\mu} = f^{\mu\nu}{}_{,\nu}$. If the antisymmetric tensor field $\sigma^{\mu\nu}$ is a solution of $\Box\sigma^{\mu\nu} = f^{\mu\nu}$ then we can set

$$\xi^{\mu} \equiv \eta^{\mu} - \sigma^{\mu\nu}{}_{,\nu}\,. \tag{2.60}$$

The vector field ξ^{μ} then satisfies the equations

$$\xi^{\mu}{}_{,\mu} = \eta^{\mu}{}_{,\mu} = \phi\,, \quad \Box\xi^{\mu} = \Box\eta^{\mu} - \Box\sigma^{\mu\nu}{}_{,\nu} = \zeta^{\mu} - f^{\mu\nu}{}_{,\nu} = 0\,. \tag{2.61}$$

Within the gauge class defined by (2.23) and (2.56) only gauge transformations which satisfy the additional conditions

$$\Box\xi^{\mu} = 0\,, \quad \xi^{\mu}{}_{,\mu} = 0 \tag{2.62}$$

are allowed. For this gauge class we have $\gamma_{\mu\nu} = h_{\mu\nu}$ as $\gamma = -h = 0$.

The most general solution of (2.55) can be written as a superposition of plane waves

$$h_{\mu\nu} = \mathrm{Re}\left(\epsilon_{\mu\nu}e^{ik_{\sigma}x^{\sigma}}\right)\,. \tag{2.63}$$

The field equations then imply

$$k^2 = k_{\mu}k^{\mu} = 0\,, \tag{2.64}$$

i.e. the wave vector of gravitational waves is null, as for light rays. The Hilbert gauge condition $\gamma^{\alpha\beta}{}_{,\beta} = h^{\alpha\beta}{}_{,\beta} = 0$ implies

$$k_{\mu}\epsilon^{\mu}_{\nu} = 0\,, \tag{2.65}$$

and the tracelessness condition (2.56) says that

$$\epsilon^{\nu}_{\nu} = 0\,. \tag{2.66}$$

The quantity $\epsilon_{\mu\nu}$ is the *polarisation tensor* of the gravitational wave, the five conditions above imply that at most five components are independent. But as we will show now, actually only two of them are independent due to residual gauge freedom.

We consider a gauge transformation induced by the vector field

$$\xi^{\mu}(x) = \mathrm{Re}\left(-i\epsilon^{\mu}e^{ik_{\sigma}x^{\sigma}}\right)\,, \tag{2.67}$$

under which the polarisation tensor changes as

$$\epsilon_{\mu\nu} \to \epsilon_{\mu\nu} + k_\mu \epsilon_\nu + k_\nu \epsilon_\mu \,. \tag{2.68}$$

For a wave propagating in the positive z direction, $k^\mu = (k, 0, 0, k)$, we then have from (2.65) that

$$\epsilon_{0\nu} = -\epsilon_{3\nu} \,. \tag{2.69}$$

This implies for example that $\epsilon_{00} = -\epsilon_{30} = -\epsilon_{03} = \epsilon_{33}$. From Eq. (2.66) we obtain $-\epsilon_{00} + \epsilon_{11} + \epsilon_{22} + \epsilon_{33} = 0$, so that

$$\epsilon_{11} + \epsilon_{22} = 0 \,. \tag{2.70}$$

We can therefore express all components of the polarisation tensor in terms of $\epsilon_{00}, \epsilon_{11}, \epsilon_{01}, \epsilon_{02}, \epsilon_{12}$. These components transform under (2.68) as

$$\epsilon_{00} \to \epsilon_{00} - 2k\epsilon_0 \,, \quad \epsilon_{11} \to \epsilon_{11} \,, \quad \epsilon_{01} \to \epsilon_{01} - k\epsilon_1 \,, \quad \epsilon_{02} \to \epsilon_{02} - k\epsilon_2 \,, \quad \epsilon_{12} \to \epsilon_{12} \,. \tag{2.71}$$

We can use ϵ_3 to ensure that ξ^μ satisfies the conditions (2.62), and the other ϵ to set all the elements of the polarisation tensor to zero, except for ϵ_{12} and $\epsilon_{11} = -\epsilon_{22}$. We therefore have only two linearly independent polarisation states, as for light.

We note that under a spatial rotation by an angle φ, e.g. about the z-axis, the polarisation tensor transforms as

$$\epsilon'_{\mu\nu} = R^\alpha{}_\mu R^\beta{}_\nu \epsilon_{\alpha\beta} \,, \tag{2.72}$$

where $R^\alpha{}_\mu$ is the usual spatial rotation matrix, with additionally $R^0{}_0 = 1$ and $R^0{}_j = R^j{}_0 = 0$. One finds

$$\epsilon'_{11} = \epsilon_{11} \cos 2\varphi + \epsilon_{12} \sin 2\varphi \,, \tag{2.73}$$
$$\epsilon'_{12} = -\epsilon_{11} \sin 2\varphi + \epsilon_{12} \cos 2\varphi, \,, \tag{2.74}$$

or for $\epsilon_\pm \equiv \epsilon_{11} \mp i\epsilon_{12}$,

$$\epsilon'_\pm = e^{\pm 2i\varphi} \epsilon_\pm \,. \tag{2.75}$$

This shows that the polarisation states $\epsilon_\pm$ have helicity (circular polarisation) ± 2, and that the polarisation tensor returns to its original state under a rotation by an angle $\varphi = \pi$. As a consequence, a quantum theory of gravity would predict massless spin-2 particles, the *gravitons*.

This symmetric tensor $\epsilon_{\mu\nu}$ where only $\epsilon_{11} = -\epsilon_{22}$ and ϵ_{12} are non-zero, is an example of a *transverse, traceless tensor*, which in general is defined by

$$h_{\mu 0} = 0 \,, \quad \sum_k h_{kk} = 0 \,, \quad h_{kj,j} = 0 \,. \tag{2.76}$$

The gauge in which a transverse, traceless tensor has this form is called the transverse, traceless (TT) gauge. In this gauge, only the components h_{ij} are non-vanishing, so that we have only six wave equations

$$\Box h_{ij} = 0 \,. \tag{2.77}$$

It is useful to compute the linearized Riemann tensor in TT gauge. As for the Ricci tensor (2.3), the linearized Riemann tensor is given by

$$R^\mu_{\sigma\nu\rho} = \partial_\nu \Gamma^\mu_{\rho\sigma} - \partial_\rho \Gamma^\mu_{\nu\sigma} \tag{2.78}$$

as the terms that are quadratic in the Christoffel symbols are of higher order. The linearized Christoffel symbols are given by (2.4), so that

$$R_{\mu\sigma\nu\rho} = g_{\mu\lambda} R^\lambda_{\sigma\nu\rho} = \frac{1}{2} \left(h_{\nu\sigma,\mu\rho} + h_{\rho\mu,\sigma\nu} - h_{\nu\mu,\sigma\rho} - h_{\rho\sigma,\mu\nu} \right) \,. \tag{2.79}$$

In TT gauge, we have that

$$R_{i0j0} = \frac{1}{2}(h_{0j,0i} + h_{i0,j0} - h_{ij,00} - h_{00,ij}) = -\frac{1}{2}h^{TT}_{ij,00}\,. \tag{2.80}$$

where we computed specifically the elements of the Riemann tensor that we will need in a moment. As the Riemann tensor produces physical, observable effects through the deviation equation, we conclude that $h_{\mu\nu}$ cannot be reduced to fewer components than what remains in the TT gauge.

Geodesic deviation by a gravitational wave

To measure gravitational waves we have to look at how they interact with matter. As gravitational waves represent a traveling, periodic distortion of space-time, they will induce oscillations of the separation of neighbouring freely falling test particles. This effect is described by the geodesic deviation equation that we studied at the end of section 1.2.7. In the weak field limit, and for bodies that are nearly 'at rest' in the global 'background' coordinate system, we can choose the vector field T^μ, which describes the geodesic of the test particles, along ∂_t, and the separation vector X^μ perpendicular and thus given by a spatial vector $\mathbf{X} = X^i$. Then the geodesic deviation equation (1.106) becomes

$$\frac{d^2X^i}{dt^2} = -R^i_{0j0}X^j\,. \tag{2.81}$$

We have, using (2.80),

$$\frac{d^2X^i}{dt^2} = \frac{1}{2}\frac{\partial^2 h^{TT}_{ij}}{\partial t^2}X^j\,. \tag{2.82}$$

Let us assume that the test-bodies were at rest relative to each other before the gravitational wave arrives, $X^i = X^i_{(0)}$ when $h_{ij} = 0$. Equation (2.82) then gives, at the location $\gamma(\tau)$ of the test particles,

$$X^i(\tau) \approx X^i_{(0)} + \frac{1}{2}h_{ij}(\gamma(\tau))X^j_{(0)}\,. \tag{2.83}$$

We consider specifically the case of a plane wave propagating (as above) in the z-direction. In the TT gauge, the non-vanishing components are

$$h^{TT}_{11} = -h^{TT}_{22} = A(t-z)\,, \quad h^{TT}_{12} = h^{TT}_{21} = B(t-z)\,. \tag{2.84}$$

Particles aligned in the z-direction, $\mathbf{X} = (0,0,a)$, see no oscillation of the their relative position, as then $h_{ij}X^j_{(0)} = 0$. Hence only *transverse oscillations* are possible. We are now going to discuss the oscillations for various polarisations. We denote the transverse components of the displacement vector $\mathbf{X}_\perp$, they satisfy the equation

$$\ddot{\mathbf{X}}_\perp \simeq K_\perp \mathbf{X}_\perp\,, \tag{2.85}$$

where

$$K_\perp = \frac{1}{2}\begin{pmatrix} \ddot{h}_{11} & \ddot{h}_{12} \\ \ddot{h}_{21} & \ddot{h}_{22} \end{pmatrix}, \quad K^T_\perp = K_\perp\,, \quad \mathrm{Tr}K_\perp = 0\,. \tag{2.86}$$

The approximate solution can then be written as

$$\mathbf{X}_\perp \simeq \mathbf{X}^{(0)}_\perp + \frac{1}{2}\begin{pmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{pmatrix}\mathbf{X}^{(0)}_\perp\,. \tag{2.87}$$

As the h-matrix is a real symmetric matrix, it can be diagonalized, and as its trace vanishes, we can write

$$\frac{1}{2}\begin{pmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{pmatrix} = R\begin{pmatrix} \Omega & 0 \\ 0 & -\Omega \end{pmatrix}R^T\,, \tag{2.88}$$

where R is an orthogonal (rotation) matrix. Introducing displacements (ξ, η) along the principal directions,

$$\mathbf{X}_\perp \equiv R \begin{pmatrix} \xi \\ \eta \end{pmatrix}, \quad \mathbf{X}_\perp^{(0)} \equiv R \begin{pmatrix} \xi_0 \\ \eta_0 \end{pmatrix}, \tag{2.89}$$

we have that

$$\xi \simeq \xi_0 + \Omega(t)\xi_0 \,, \quad \eta \simeq \eta_0 - \Omega(t)\eta_0 \,. \tag{2.90}$$

As the principal directions are orthogonal, we find a "quadrupole" type oscillation pattern about the point (ξ_0, η_0).

For a periodic plane wave propagating in the z-direction the functions A and B of (2.84) can be written as

$$h_{11} = -h_{22} = \mathrm{Re}\left[A_+ e^{-i\omega(t-z)}\right], \quad h_{12} = h_{21} = \mathrm{Re}\left[A_\times e^{-i\omega(t-z)}\right]. \tag{2.91}$$

If $A_\times = 0$ then the principal directions are aligned with the x- and y-axis, while if $A_+ = 0$ then they are rotated by $45°$. In these two cases we call the wave linearly polarized. For $A_\times = \pm i A_+$ the wave is said to be (right, resp. left) circularly polarized.

Gravitational wave detectors

Gravitational wave detectors are instruments built to detect precisely the oscillations of test-masses due to space-time deformation of a passing gravitational wave, as described above. The advanced LIGO / Virgo detectors that measured the first gravitational waves passing through Earth are large-scale interferometers. Here we will consider a simple toy-detector built from two masses m that are connected by massless spring of equilibrium length $2l_0$. The vector connecting the two masses forms an angle ϑ with the z-axis and φ with the x-axis. Let ω_0 be the frequency of the oscillator, and we assume that the damping time $\tau \ll 1/\omega_0$. We write the separation of the masses as $2(l_0 + \xi(t))$, then the relative separation $\xi(t)$ will satisfy the differential equation

$$\ddot{\xi} + \frac{1}{\tau_0}\dot{\xi} + \omega_0^2 \xi = \text{tidal acceleration} \,. \tag{2.92}$$

The tidal acceleration is given by our results above. We assume an incoming periodic gravitational wave described by (2.91), with a wave length much larger than the size of the detector. Choosing the polarisation aligned with x- and y-axis ($A_\times = 0$), the tidal acceleration matrix $K_\perp$ in (2.86) is

$$K_\perp = -\frac{1}{2}\omega^2 \mathrm{Re}\left(A_+ e^{-i\omega t}\right) \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{2.93}$$

The tidal acceleration of the detector masses is then

$$-\frac{1}{2}\omega^2 \mathrm{Re}\left(A_+ e^{-i\omega t}\right) l_0 \left(\sin^2\vartheta \cos^2\varphi - \sin^2\vartheta \sin^2\varphi\right), \tag{2.94}$$

and the driven oscillator equation (2.92) becomes

$$\ddot{\xi} + \frac{1}{\tau_0}\dot{\xi} + \omega_0^2 \xi = -\frac{1}{2}\omega^2 l_0 \mathrm{Re}\left(A_+ e^{-i\omega t}\right) \sin^2\vartheta \cos 2\varphi \,. \tag{2.95}$$

The dependence on 2φ reflects once more the spin-2 nature of gravitational waves, and the factor $\sin^2\vartheta$ is due to their transverse nature. As you can see, the strength of a received GW signal depends on the relative orientation of the detector. On the other hand, with several detectors it is possible to determine the polarisation and the propagation direction of the wave due to the angular dependence.

The driven solution then is

$$\xi = \mathrm{Re}\left(\mathcal{A}_+ e^{-i\omega t}\right) , \quad \mathcal{A}_+ = \frac{1}{2} A_+ l_0 \sin^2\vartheta \cos 2\varphi \frac{\omega^2}{\omega^2 - \omega_0^2 + i\omega/\tau_0} . \tag{2.96}$$

The response of the detector in terms of $\xi/l_0 \sim \Delta l/l$ is thus proportional to the incoming GW amplitude A_+, the angular 'acceptance factor' $\sin^2\vartheta\cos 2\varphi$ that depends on the wave polarisation and propagation direction, and a frequency dependent response factor of the detector system.

The instantaneous energy of the oscillator is given by the kinetic and potential contributions,

$$E = 2\left(\frac{1}{2} m\dot{\xi}^2 + \frac{1}{2} m\omega_0 \xi^2\right) . \tag{2.97}$$

Averaging this over one period of the driven solution we get

$$\langle E \rangle = \frac{1}{2} m |\mathcal{A}_+|^2 (\omega^2 + \omega_0^2) = \frac{1}{8} m l_0^2 A_+^2 \sin^4\vartheta \cos^2 2\varphi \frac{\omega^4(\omega^2+\omega_0^2)}{(\omega^2-\omega_0^2)^2 + \omega^2/\tau_0^2} . \tag{2.98}$$

The signal, as a function of the GW frequency ω, is maximal at the resonance frequency of the oscillator, $\omega \approx \omega_0$. Defining the 'Q' (quality factor) of the oscillator as $Q \equiv \omega_0 \tau_0$ we have

$$\langle E \rangle_{\mathrm{res}} = \frac{1}{4} m l_0^2 \omega_0^2 A_+^2 \sin^4\vartheta \cos^2 2\varphi \, Q^2 . \tag{2.99}$$

The LIGO detector[4] does not involve test masses connected by a spring (our toy model is instead more similar to resonant bars). Instead it uses test masses on damped and environmentally isolated pendulums as shown schematically in Fig. 2.2. The arm length of LIGO is 4km, and it contains two orthogonal arms. A laser interferometer measures whether the distance of the perpendicular arms change – due to the quadrupole deformation pattern of a gravitational wave, the orthogonal arms produce the maximal effect. The interference of the two light beams is adjusted so that they interfere destructively at a photo detector, so that the detector sees no light if the test masses are at rest. If a GW arrives, the arm lengths will change and the destructive interference will only be partial. Of course in order to reach the required sensitivity the advanced LIGO and Virgo detectors use many additional tricks to increase the laser power and reduce the noise.

From (2.87) and (2.91) we see that the relative amplitude $\Delta l/l$ of spatial oscillations between test particles induced by a linearly polarized harmonic gravitational plane wave is equal to $A_+/2$ or $A_\times/2$. This motivates the definition for the total *strain* of the space-time deformation by a gravitational wave,

$$h \equiv \sqrt{A_+^2 + A_\times^2} \sim 2\frac{\Delta l}{l} . \tag{2.100}$$

Figure 2.3 shows the sensitivity in terms of the strain detectable by LIGO and other detectors (briefly described below). Advanced LIGO is sensitive to strains of about 10^{-22} in its frequency band. This means that it can detect a length change of $\Delta l \approx 10^{-22} l \approx 4 \times 10^{-19}$ m, which is over 10'000 times smaller than the charge radius of a proton (the 'size' of a proton that an electron sees in scattering experiments)! The figure also shows the expected strain at Earth from some sources, a topic that we will explore a bit further in the next section.

This amazing sensitivity enabled the first direct detection of gravitational waves at 9:51 UTC on September 14, 2015, from a merger of two black holes of about 30 solar masses each, at a distance of about 1.3 billion light years from Earth. The measured signal is shown in Fig. 2.4. A later detection, on

[4]http://www.ligo.org/

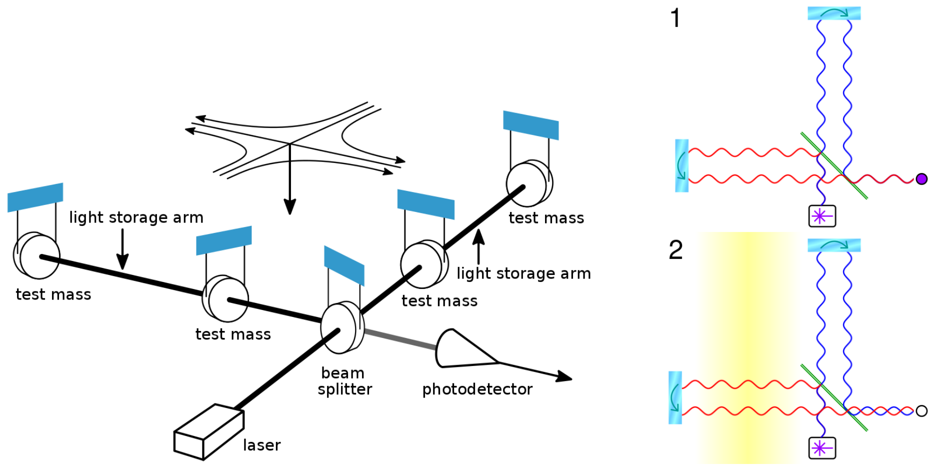

Figure 2.2: Left: Schematic of the (advanced) LIGO detector set up. Right: Illustration of the detection principle. The light from the laser enters the two arms through a beam splitter. After reflection at the arm ends, the light is recombined and interferes. Any change in the interference pattern due to changes in relative arm lengths is detected by the photodetector. Figures from Wikipedia.

August 17th 2017, due to the merger of two neutron stars, was accompanied by a gamma ray burst (GRB) and an optical afterglow. The time delay between the arrival of the GW signal and the GRB places an upper limit of

$$\left|\frac{c_T - c}{c}\right| \lesssim 10^{-15} \tag{2.101}$$

on the relative difference of the speed of light c and the speed of gravitational waves c_T. These observations represent (yet another) spectacular confirmation of the predictions of General Relativity, and impose strong constraints on any alternative theories. The 2017 Nobel prize in physics was awarded to Rainer Weiss, Kip Thorne and Barry Barish in recognition of their work that made these discoveries possible.

Detectors on Earth are invariably exposed to many sources of noise, geological and of human origin that limit the range of frequencies at which GW measurements are possible. In addition, arm lengths are limited, but longer arms would give a larger sensitivity. For these reasons there are plans to place a gravitational wave detector in space. Thanks to the boost that the field received with the recent discoveries of GW's, the European Space Agency has selected LISA[5] as the 'L3' mission, with a launch date expected currently near the retirement age of your lecturer.

Another approach to measure gravitational wave is to use radio signals from pulsars (a rotating, strongly magnetized neutron star that emits a narrow radio beam which periodically scans across Earth and so creates a periodic signal) as highly precise clocks. If gravitational waves pass between us the pulsars, the light propagation time is modulated. By monitoring simultaneously large numbers of pulsars, the sensitivity can be greatly increased. The lower frequency limit is given by the duration of the observation, after thirty years it is of the order of nano-Hertz.

2.4 Emission of gravitational waves

In order to be able to estimate the strength of a gravitational wave signal on Earth we need to develop at least a basic idea of how gravitational waves are emitted by astrophysical objects. To study the emission

[5] https://www.lisamission.org/

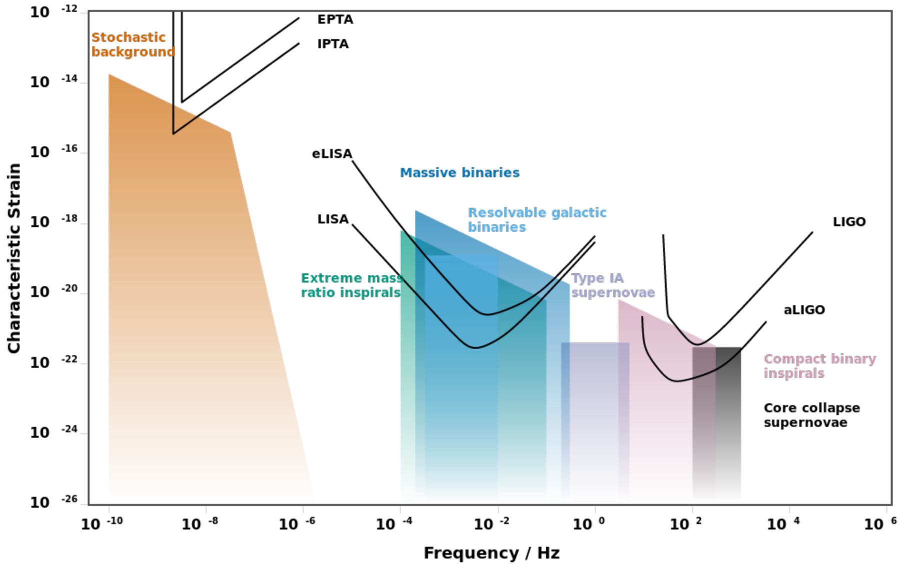

Figure 2.3: Detector noise curves for Initial and Advanced LIGO as a function of frequency. LIGO is sensitive at frequencies of about ten to thousand Hertz. At lower frequencies (inaccessible on Earth due to geological noise) lie the bands for space-borne detectors like the Laser Interferometer Space Antenna (LISA) and pulsar timing arrays such as the European Pulsar Timing Array (EPTA). The characteristic strains of potential astrophysical sources are also shown. To be detectable the characteristic strain of a signal must be above the noise curve. Image from Wikipedia (authors: Christopher Moore, Robert Cole and Christopher Berry, http://rhcole.com/apps/GWplotter/).

of gravitational waves, we return to the retarded solution (2.30) in Hilbert gauge,

$$\gamma_{\mu\nu}(x) = 4G \int \frac{T_{\mu\nu}(x^0 - |\mathbf{x}-\mathbf{y}|, \mathbf{y})}{|\mathbf{x}-\mathbf{y}|} d^3y \,. \tag{2.102}$$

At large distances from a localized source we can replace the denominator of this equation simply by $r \equiv |\mathbf{x}|$. The retarded time can be approximated by, writing $t \equiv x^0$,

$$t - |\mathbf{x}-\mathbf{y}| \simeq t - r + \mathbf{y}\cdot\hat{\mathbf{x}} \,, \quad \hat{\mathbf{x}} \equiv \mathbf{x}/r \,. \tag{2.103}$$

The asymptotic metric then becomes

$$\gamma^{\mu\nu}(t,\mathbf{x}) = \frac{4G}{r} \int T^{\mu\nu}(t - r + \mathbf{y}\cdot\hat{\mathbf{x}}, \mathbf{y}) d^3\mathbf{y} \,. \tag{2.104}$$

The equation above can be used for arbitrarily fast time variations of $T^{\mu\nu}$, but let's assume that motions inside the source are sufficiently slow so that we can approximate the retarded time by $t - r$,

$$\gamma^{\mu\nu}(t,\mathbf{x}) = \frac{4G}{r} \int T^{\mu\nu}(t - r, \mathbf{y}) d^3\mathbf{y} \,. \tag{2.105}$$

To compute the perturbed metric in TT gauge we only need the space-space components, but the stress-energy density of the source will be dominated by the time-time component. Thanks to the linearized conservation law ${T^{\mu\nu}}_{,\nu} = 0$ we can connect the two. We start from

$$0 = \int x^k \partial_\nu T^\nu_\mu d^3x = \frac{\partial}{\partial t} \int x^k T^0_\mu d^3x + \int x^k \partial_j T^j_\mu d^3x \,. \tag{2.106}$$

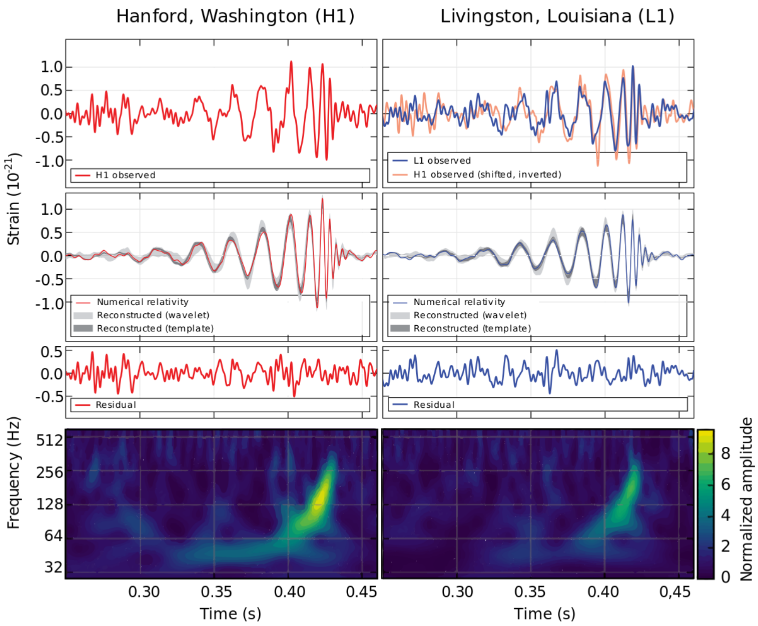

Figure 2.4: The observed strain (top panel), best-fit template (second from top), residual and time-frequency plot for the first GW signal directly detected on Earth, GW150914. The left and right plots show the signals at the two advanced LIGO detectors (Hanford and Livingston). The signal was so strong that it is clearly visible over the (filtered) noise. About 3 solar masses were turned into GW energy and emitted during the merger of the two black holes. Reference: Abbott, Benjamin P.; et al. (LIGO Scientific Collaboration and Virgo Collaboration), *Observation of Gravitational Waves from a Binary Black Hole Merger*, Phys. Rev. Lett. **116**, 061102 (2016), arXiv:1602.03837.

Integrating the last expression by parts we obtain

$$\int T^k_\mu d^3x = \frac{\partial}{\partial t} \int T^0_\mu x^k d^3x \,. \tag{2.107}$$

In addition, we can write (using the conservation of $T^{\mu\nu}$ and partial integration)

$$\frac{\partial}{\partial t} \int T^{00} x^k x^j d^3x = -\int \left(\partial_l T^{l0}\right) x^k x^j d^3x = \int T^{l0} \partial_l \left(x^k x^j\right) d^3x = \int \left(T^{k0} x^j + T^{j0} x^k\right) d^3x \,. \tag{2.108}$$

Combining the last two equations we finally find

$$\frac{1}{2} \frac{\partial^2}{\partial t^2} \int T^{00} x^k x^j d^3x = \frac{1}{2} \frac{\partial}{\partial t} \int \left(T^{k0} x^j + T^{j0} x^k\right) d^3x = \int T^{kj} d^3x \,. \tag{2.109}$$

Considering only nearly-Newtonian sources for which the energy density is dominated by the matter density ρ, i.e. $T^{00} \sim \rho$, we then have from (2.105) and (2.109)

$$\gamma^{kj} \simeq \frac{2G}{r} \frac{\partial^2}{\partial t^2} \int \rho(t - r, \mathbf{y}) y^k y^j d^3y \,. \tag{2.110}$$

Introducing the trace-free *quadrupole tensor* (notice that normalisation conventions can differ)

$$Q_{kj}(t) = \int (3x_k x_j - r^2 \delta_{kj}) \rho(t, \mathbf{x}) d^3x \tag{2.111}$$

we can rewrite the equation for γ^{kj} as

$$\gamma_{kj}(t,\mathbf{x}) = \frac{2G}{3}\frac{1}{r}\left(\frac{\partial^2}{\partial t^2}Q_{kj}(t-r) + \delta_{kj}\frac{\partial^2}{\partial t^2}\int r'^2\rho(t-r,\mathbf{y})d^3y\right) . \tag{2.112}$$

In the TT gauge this gives

$$h^{TT}_{jk}(t,\mathbf{x}) = \frac{2G}{3}\frac{1}{r}\ddot{Q}^{TT}_{jk}(t-|\mathbf{x}|) . \tag{2.113}$$

This expression tells us what the metric perturbation is that gives rise to a gravitational wave, at large distances from the source. The lowest multipole emitting gravitational waves is the quadrupole, in contrast to electrodynamics where it is the dipole. This is due to the spin-2 nature of the gravitational waves, i.e. the fact that they are described by a rank-2 tensor.

Energy emitted in the form of gravitational waves

Computing the energy flux of a gravitational wave is relatively involved. Details can be found for example in section 5.4 of [2], here we just use the results. For a wave that can locally be considered a plane wave (certainly a good approximation far from the source), the energy flux in e.g. the x^1-direction is

$$T^{01}_{(\mathrm{GW})} = \frac{1}{32\pi G}\left\langle 2(\dot{\gamma}_{23})^2 + \frac{1}{2}(\dot{\gamma}_{22} - \dot{\gamma}_{33})^2\right\rangle . \tag{2.114}$$

The angle brackets here denote time averages over several characteristic periods of the source. From the equation for a plane wave (2.91) and the definition of the strain (2.100) we note that we can write the flux as

$$\mathcal{F}_{\mathrm{GW}} = \frac{1}{32\pi G}\omega^2(A_+^2 + A_\times^2) = \frac{1}{32\pi G}\omega^2 h^2 . \tag{2.115}$$

From our solution for the metric (2.112) the expression for $T^{01}_{(\mathrm{GW})}$ is

$$T^{01}_{(\mathrm{GW})} = \frac{G}{72\pi}\frac{1}{r^2}\left\langle 2(\dddot{Q}_{23})^2 + \frac{1}{2}(\dddot{Q}_{22} - \dddot{Q}_{33})^2\right\rangle . \tag{2.116}$$

This can actually be written in terms of the transverse, trace-free part of the quadrupole tensor,

$$T^{0s}_{(\mathrm{GW})}n^s = \frac{G}{72\pi}\frac{1}{r^2}\left\langle \mathrm{Tr}\left(\dddot{Q}^{TT}\right)^2\right\rangle , \tag{2.117}$$

where

$$\mathrm{Tr}\left(\dddot{Q}^{TT}\right)^2 = \dddot{Q}_{kl}\dddot{Q}_{kl} - 2\dddot{Q}_{kl}\dddot{Q}_{km}n^l n^m + \frac{1}{2}\left(\dddot{Q}_{kl}n^k n^l\right)^2 . \tag{2.118}$$

The average power per unit solid angle, emitted in the direction $\mathbf{n}$, is then given by (2.117) times r^2 (from the surface volume element),

$$\frac{dL_{\mathrm{GW}}}{d\Omega} = \frac{G}{72\pi}\left\langle \mathrm{Tr}\left(\dddot{Q}^{TT}\right)^2\right\rangle . \tag{2.119}$$

The total power (gravitational wave luminosity) is the integral over the sky sphere. Using (2.117) and the averages (similar to the relevant computations in electrodynamics),

$$\frac{1}{4\pi}\int_{S^2} n^i n^j d\Omega = \frac{1}{3}\delta_{ij} , \tag{2.120}$$

$$\frac{1}{4\pi}\int_{S^2} n^i n^j n^k n^l d\Omega = \frac{1}{15}\left(\delta_{ij}\delta_{kl} + \delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}\right) , \tag{2.121}$$

we find the famous *quadrupole formula* of Einstein,

$$L_{\mathrm{GW}} = \frac{G}{45}\left\langle \dddot{Q}_{ij}\dddot{Q}_{ij}\right\rangle . \tag{2.122}$$

Emission from a rigidly rotating body

The simplest example in which we can compute the emission of gravitational waves is a rigidly rotating body, an example that includes binary stars in circular orbits about the common center of mass.

We assume that a body with mass distribution $\rho(\mathbf{x}')$ (in fixed Euclidean coordinates) is rotating with angular frequency ω about the z-axis. Then in the inertial (observer / laboratory) frame the mass distribution is $\rho(\mathbf{x}, t) = \rho(\mathbf{x}')$, where

$$\begin{aligned} x_1 &= x'_1 \cos\omega t - x'_2 \sin\omega t \\ x_2 &= x'_1 \sin\omega t + x'_2 \cos\omega t \\ x_3 &= x'_3 \,. \end{aligned} \tag{2.123}$$

The moment-of-inertia tensor in fixed body coordinates is given by

$$I'_{ij} = \int \rho(\mathbf{x}') x'_i x'_j d^3x' \,. \tag{2.124}$$

For simplicity we will assume that it is diagonal in our coordinate system (and thus the rotation axis is aligned with one of the principal axis of the inertia tensor, and the other principal axes are also oriented along coordinate axes). In the inertial observer system the components of I then are

$$\begin{aligned} I_{11} &= \frac{1}{2}\left(I'_{11} + I'_{22}\right) + \frac{1}{2}\left(I'_{11} - I'_{22}\right)\cos 2\omega t \,, \\ I_{12} &= \frac{1}{2}\left(I'_{11} - I'_{22}\right)\sin 2\omega t \,, \\ I_{22} &= \frac{1}{2}\left(I'_{11} + I'_{22}\right) - \frac{1}{2}\left(I'_{11} - I'_{22}\right)\cos 2\omega t \,, \\ I_{13} &= I_{23} = 0 \,, \\ I_{33} &= I'_{33} \,. \end{aligned} \tag{2.125}$$

We see that the time dependence is twice the rotation frequency, as expected for a rotating ellipsoid. The trace-free quadrupole tensor is given by

$$Q_{ij} = 3I_{ij} - \delta_{ij} I_{kk} \tag{2.126}$$

which determines h_{ij}^{TT} through (2.113) and L_{GW} through (2.122). We find that the metric at a distance D in the 1-direction is given by

$$h_{22}^{TT} = -h_{33}^{TT} = \frac{1}{2}(h_{22} - h_{33}) = \frac{1}{2} G\theta e (2\omega)^2 \frac{1}{D} \cos 2\omega t \,, \quad h_{23}^{TT} = h_{23} = 0 \,. \tag{2.127}$$

where

$$\theta = I'_{11} + I'_{22} \,, \quad e = \frac{I'_{11} - I'_{22}}{\theta} \,. \tag{2.128}$$

This corresponds to a linearly polarized wave with

$$A_+ = h = \frac{2}{D} G\theta e \omega^2 \,, \tag{2.129}$$

while along the 3-axis the wave is circularly polarized, with $A_\times = iA_+$ and

$$A_+ = \frac{4}{D} G\theta e \omega^2 \,, \tag{2.130}$$

giving a total strain of

$$h = \frac{4\sqrt{2}}{D} G\theta e\omega^2 \,. \tag{2.131}$$

We see that a measurement of the polarisation of gravitational waves can give information on the inclination of the orbit of binary neutron stars or black holes.

The quadrupole formula (2.122) gives for the total power (after a bit of calculation)

$$L_{\rm GW} = \frac{32G}{5c^2}\theta^2 e^2 \omega^6 \,, \tag{2.132}$$

where we included dimensionful constants. As a simple example, for a homogeneous rotating rod of length l and mass M we have $\theta = (Ml^2)/12$ and $e = 1$, so that

$$L_{\rm GW} = \frac{2}{45}\frac{G}{c^2} M^2 l^4 \omega^6 \,. \tag{2.133}$$

For an iron rod with $l = 100$m, $m = 1000$ tons, and a rotation frequency of 3 Hz, the total power of GW emission is $L_{\rm GW} \simeq 10^{-26}$W, which is very small.

A more relevant example is a binary star system of masses m_1 and m_2 in circular orbits around their common center of mass. If r is the separation of the two orbits, then $e = 1, \theta = \mu r^2$ where μ is the reduced mass, $\mu = m_1 m_2/(m_1 + m_2)$. From Kepler's third law we have that

$$\omega^2 = \frac{GM}{r^3} \,, \quad M = m_1 + m_2 \,. \tag{2.134}$$

We then find from (2.132),

$$L_{\rm GW} = \frac{32}{5}\frac{G^4}{c^5 r^5} M^3 \mu^2 = \frac{32}{5}\left(\frac{GM}{c^2 r}\right)^5 \frac{\mu^2}{M^2} L_0 \,, \tag{2.135}$$

where

$$L_0 \equiv \frac{c^5}{G} = 3.63 \times 10^{59} \text{erg/sec} = 3.63 \times 10^{52} \text{W} = 2.03 \times 10^5 M_\odot c^2/\text{sec} \,. \tag{2.136}$$

In the end-stadium of an inspiral the emitted power in gravitational waves can become huge. For example, two neutron stars with masses of about $1.4 M_\odot$ at a separation of $r = 100$km and corresponding orbital period of about 10^{-2}s, then the luminosity from (2.135) is of the order of $L_{\rm GW} \simeq 10^{45}$W. The peak luminosity during the merger of the $\sim 30 M_\odot$ black holes that led to the first detection of gravitational waves, GW150914, was about 3.6×10^{49}W, more than the power of all light radiated by stars in the observable universe. In that event, an energy of about $3M_\odot$ was radiated away as GW during the inspiral and merger.

An order of magnitude estimate for the emission of gravitational waves from a system of size R of orbiting masses M is thus

$$L_{\rm GW} \sim L_0 \left(\frac{GM}{c^2 R}\right)^5 \,. \tag{2.137}$$

As the Schwarzschild radius of an object with mass M is $R_S = 2GM/c^2$, the 'compactness parameter' is usually very small. But for two coalescing black holes it can become of order unity at the end of an inspiraling phase, leading to the huge output of gravitational wave power mentioned above. This huge emitted power, together with the high sensitivity of the advanced interferometric detectors, enables the direct detection of gravitational waves from distant sources. It should be noted that such gravitational waves are probably the only practicable way to directly observe black holes.

The total strain (2.131) for such a binary star system will depend on the (luminosity) distance D and the output power. It is, using also (2.134),

$$h = 4\sqrt{2}\frac{1}{D}\frac{G^2 M\mu}{r} = 4\sqrt{2}\frac{Gm_1/c^2}{D}\frac{Gm_2/c^2}{r} = 4\sqrt{2}\frac{1}{D}(G\mathcal{M})^{5/3}\omega^{2/3}\,, \tag{2.138}$$

where we introduced the *chirp mass*

$$\mathcal{M} \equiv \frac{(m_1 m_2)^{3/5}}{(m_1+m_2)^{1/5}} = \mu^{3/5}M^{2/5}\,. \tag{2.139}$$

Using both the amplitude and the time evolution of the signal it is possible to determine separately D and $\mathcal{M}$.

In the above example of the two neutron stars, we find for $D \approx 30'000$ light years (the distance to the center of the Milky Way) a total strain of $h \approx 10^{-18}$. This would leave a huge signal in a laser interferometer with a strain sensitivity of about 10^{-22} at frequencies of tens of Hertz (cf Fig. 2.3). The black hole merger GW150914 was at a distance of about 440Mpc (1.4 billion light years), and the peak observed strain was about 10^{-21} as can be seen in Fig. 2.4.

To predict how often we would be able to observe events with a given strain requires not only the computation of h for given systems, but also an idea of how often such systems exist in a given volume, and hence what the probability of a GW emission event per unit time and as a function of D is. This is an astrophysical question that goes beyond the scope of this lecture – the large masses of the black holes involved in the merger that produced the first GW detection with advanced LIGO was certainly a surprise for astronomers. In the future GW astronomy may become crucial for a better understanding of the formation and evolution of stellar systems and their remnants.

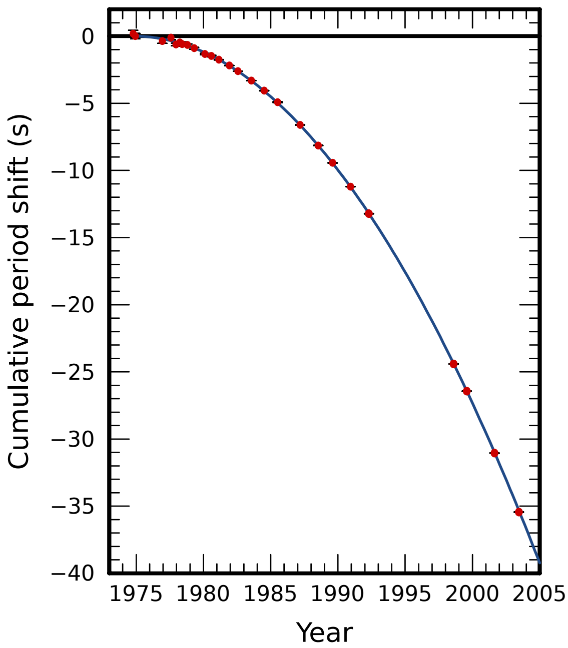

Figure 2.5: Evolution of the pulse arrival period of the binary pulsar PSR B1913+16, also knowns as 'Hulse-Taylor binary'. The speed-up is clearly visible, and the prediction from GR (solid line) fits the observations perfectly. Based on this data, the final inspiral should happen in about 300 million years. (Graph from Wikipedia based on data from Weisberg & Taylor.)

While GW150914 was the first direct detection of a gravitational wave, there was strong observational evidence for their existence already beforehand. In 1974 Russell Alan Hulse and Joseph Hooton Taylor, Jr., discovered a pulsar in a binary system. Over several decades the measured the change in orbital period, which was interpreted as being due to energy loss from the emission of gravitational waves. Figure 2.5 shows the data and the prediction from General Relativity. Hulse & Taylor received the Nobel prize in physics in 1993 for their work.

2.5 Gravitational lensing

Light travels on null geodesics, which in general are not straight lines but are curved if the metric is non-trivial. This changes the apparent positions of light sources as seen by an observer, and induces magnification and image distortions ('shear'). For this reason, gravitational lensing can be used to derive information about the metric perturbations from astronomical observations. This astronomical technique has become more and more important over the last years, and it will continue to do so over the next decade as several large astronomical surveys are based on gravitational lensing.

2.5.1 Lensing and time delay in nearly-Newtonian situations

We now use our results for nearly-Newtonian fields to study the propagation of light rays. We have seen that the metric in this limit is given by (2.34) for the 'Newtonian' gravitational potential ϕ

$$\Delta\phi = 4\pi G\rho\,, \tag{2.140}$$

and that photons follow null geodesics, Eq. (1.195). The linearized Christoffel symbols (2.4) of the nearly-Newtonian metric are

$$\Gamma^0_{0i} = \Gamma^0_{i0} = \Gamma^i_{00} = \partial_i\phi\,, \quad \Gamma^i_{jk} = \delta_{jk}\partial_i\phi - \delta_{ik}\partial_j\phi - \delta_{ij}\partial_k\phi\,. \tag{2.141}$$

We write the photon trajectory as $x^\mu(\lambda)$. Following [4] we split the geodesic into a 'background' path and a perturbation,

$$x^\mu(\lambda) = x^{(0)\mu}(\lambda) + x^{(1)\mu}(\lambda)\,. \tag{2.142}$$

The background photon trajectory $x^{(0)\mu}(\lambda)$ describes the photon propagation in the unperturbed background metric $\eta_{\mu\nu}$ and will thus just be a straight null path (a null geodesic of the flat background metric). The vector $x^{(1)\mu}(\lambda)$ then describes the deviation between the background and the true photon trajectory. As we only work to first order in perturbations, we can evaluate all perturbed quantities along that background path to find $x^{(1)\mu}(\lambda)$. This works as long as the metric along the background and the perturbed path is not too different, corresponding to the condition $x^{(1)i}\partial_i\phi \ll \phi$. We can always improve our approximation by considering only short segments, where the difference between the unperturbed and the perturbed trajectory will necessarily be small, and then assembling the full trajectory from such segments.

The conditions on the photon trajectory (to be solved order by order) are that the trajectory is null,

$$g_{\mu\nu}\frac{dx^\mu}{d\lambda}\frac{dx^\nu}{d\lambda} = 0\,, \tag{2.143}$$

and a geodesic,

$$\frac{d^2x^\mu}{d\lambda^2} + \Gamma^\mu_{\rho\sigma}\frac{dx^\rho}{d\lambda}\frac{dx^\sigma}{d\lambda} = 0\,. \tag{2.144}$$

We denote the background wave vector as $k^\mu = dx^{(0)\mu}/d\lambda$ and the derivative of the deviation vector as $\ell^\mu = dx^{(1)\mu}/d\lambda$.

The null condition for the background path gives $\eta_{\mu\nu}k^\mu k^\nu = 0$, or $(k^0)^2 = \mathbf{k}^2 \equiv k^2$. At first order we find

$$2\eta_{\mu\nu}k^\mu \ell^\nu + h_{\mu\nu}k^\mu k^\nu = 0\,, \tag{2.145}$$

or

$$-k\ell^0 + \mathbf{k}\cdot\boldsymbol{\ell} = 2k^2\phi\,. \tag{2.146}$$

The geodesic equation at zeroth order, $d^2x^{(0)\mu}/d\lambda^2 = 0$, just tells us that $x^{(0)\mu}$ is a straight path, while at first order we obtain

$$\frac{d\ell^\mu}{d\lambda} = -\Gamma^\mu_{\rho\sigma}k^\rho k^\sigma\,. \tag{2.147}$$

In terms of temporal and spatial components we have

$$\frac{d\ell^0}{d\lambda} = -2k(\mathbf{k}\cdot\nabla\phi)\,, \quad \frac{d\boldsymbol{\ell}}{d\lambda} = -2k^2\nabla_\perp\phi\,. \tag{2.148}$$

Here we introduced the derivative perpendicular to the propagation direction, defined as the total gradient minus the gradient along the path,

$$\nabla_\perp\phi \equiv \nabla\phi - \nabla_\parallel\phi = \nabla\phi - (\hat{\mathbf{k}}\cdot\nabla\phi)\hat{\mathbf{k}}\,, \tag{2.149}$$

for $\hat{\mathbf{k}} \equiv \mathbf{k}/k$ the unit vector in the $\mathbf{k}$ direction.

We can integrate the temporal geodesic equation, with the additional condition that $\ell^0 = 0$ for $\phi = 0$ to fix the constant of integration,

$$\ell^0 = \int \frac{d\ell^0}{d\lambda}d\lambda = -2k\int(\mathbf{k}\cdot\nabla\phi)d\lambda = -2k\int\left(\frac{d\mathbf{x}}{d\lambda}\cdot\nabla\phi\right)d\lambda = -2k\int\nabla\phi d\mathbf{x} = -2k\phi\,. \tag{2.150}$$

Inserting this into the null condition (2.146) we find

$$\boldsymbol{\ell}\cdot\mathbf{k} = k\ell^0 + 2k^2\phi = 0\,, \tag{2.151}$$

which shows that $\boldsymbol{\ell}$ and $\mathbf{k}$ are perpendicular to first order.

The spatial part of the geodesic equation shows that rate of change of the deviation vector $\boldsymbol{\ell}$ along the photon trajectory is given by the perpendicular gradient of the gravitational potential, and from the Poisson equation (2.140) we see that mass concentrations generate the gravitational potential. We thus expect mass concentrations near the photon path to bend the light ray, and we introduce the *deflection angle* $\hat{\alpha}$ to describe this effect.We define the deflection angle as rotation of the wave vector between the beginning and the end of the trajectory,

$$\hat{\alpha} = -\frac{\Delta\boldsymbol{\ell}}{k}\,, \quad \Delta\boldsymbol{\ell} = \int\frac{d\boldsymbol{\ell}}{d\lambda}d\lambda = -2k^2\int\nabla_\perp\phi\,d\lambda\,. \tag{2.152}$$

The deflection angle is a two-dimensional vector in the plane perpendicular to $\mathbf{k}$, the minus sign is due to it being measured by an observer looking back along the photon path. Expressing the deflection angle as an integral over physical spatial distance, $s = k\lambda$, we find the simple expression

$$\hat{\alpha} = 2\int\nabla_\perp\phi\,ds\,. \tag{2.153}$$

In addition to the deflection of light there is also a gravitational time delay, often called *Shapiro time delay*. The total coordinate time elapsed along a null path is

$$t = \int\frac{dx^0}{d\lambda}d\lambda\,. \tag{2.154}$$

Assuming that the observer is located far from any source of gravity, and at rest in the inertial background frame, we can consider the coordinate time to be the proper time of the observer. The presence of a Newtonian potential along the photon path then 'slows down' the photons, relative to the situation without that potential, and gives a time delay of

$$\Delta t = \int \frac{dx^{(1)0}}{d\lambda} d\lambda = \int \ell^0 d\lambda = -2k \int \phi d\lambda \, , \tag{2.155}$$

or expressed again in terms of an integral over physical spatial distance s,

$$\Delta t = -2 \int \phi ds \, . \tag{2.156}$$

In our perturbative setting we can perform the integral over the background path. This time delay was first proposed by Irwin Shapiro in 1964 and observed a few years later by comparing the travel time of radar signals between Earth and Venus, with and without the sun nearby the path. The time difference in that case is of about 200 μs.

It should be noted that there is additionally a 'geometric' time delay due to the longer path photons have to travel along the curved path, relative to the unlensed background trajectory. For the solar system observation above this geometric time delay is negligible, but in cosmological settings it can be of the same order of magnitude as Shapiro time delay.

The predictions of General Relativity for both the deflection of light (gravitational lensing) and the Shapiro time delay have been tested with observations. The most famous example is probably the original expedition by Eddington and colleagues during a solar eclipse in 1919, where the prediction of GR was confirmed, a critical milestone for the adoption of GR as the correct description of gravity. There was actually an earlier German expedition to observe a solar eclipse in the Crimean in 1914, but the outbreak of the war stopped this effort (which was just as well, as Einstein's predictions at the time were wrong). Figure 2.6 shows how the constraints over time have improved, and how the results are in excellent agreement with GR predictions.

2.5.2 Lensing by a point mass

The lensing of light by the gravitational field of the sun was an important historical test of GR. While GR also explained the perihelion precession of Mercury, this was a discrepancy that was already known at the time. The light deflection by the sun on the other hand had not yet been observed, and hence GR made a pure prediction. It was measured for the first time in a famous expedition led by Eddington in 1919, during a total solar eclipse. However, as we will see the effect is very small for the sun, and it is not certain if it could have been measured with the precision claimed at the time. At any rate, today the agreement with GR is very well established.

We assume that a light ray propagating in the x direction passes by a point mass M, with an impact parameter (minimal distance) b. The gravitational potential along the light ray is then given by

$$\phi = -\frac{GM}{r} = -\frac{GM}{\sqrt{b^2 + x^2}} \, . \tag{2.157}$$

The transverse gradient is then

$$\nabla_\perp \phi = \frac{GM}{(b^2 + x^2)^{3/2}} \mathbf{b} \, , \tag{2.158}$$

with the vector $\mathbf{b}$ being the transverse vector of length b from the photon path to the point source at the smallest distance between the light ray and the mass. The deflection is

$$\hat{\alpha} = 2GMb \int \frac{dx}{(b^2 + x^2)^{3/2}} = \frac{4GM}{b} \, , \tag{2.159}$$

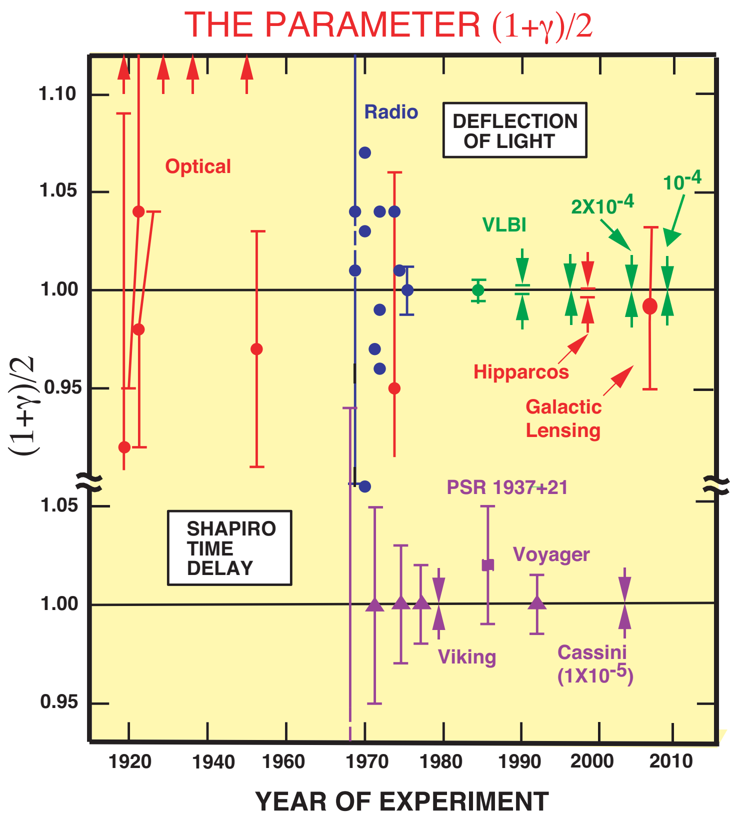

Figure 2.6: Comparison between the predictions of GR and observations, for gravitational lensing (upper part) and time delay (lower part). From [7].

where we integrated the path from $-\infty$ to ∞, assuming that the light source and the observer are sufficiently far from the point mass to make this a good approximation.

For the sun, $GM_\odot/c^2 = 1.48 \times 10^3$m (note that we set $c = 1$ in the expressions above) and $R_\odot = 6.96 \times 10^8$m, so that the maximal deflection angle (for $b = R_\odot$) is $\hat{\alpha} = 1.75$ arc sec. This is not much larger than the smearing of astronomical images due to turbulence in the atmosphere ('astronomical seeing'), which is of the order of 0.4 arc sec or more.

2.5.3 Lensing by arbitrary mass distributions

We can write (2.153) as

$$\nabla_\perp \cdot \hat{\alpha} = 2 \int \Delta_\perp \phi ds \,. \tag{2.160}$$

We assume that we can replace the two-dimensional Laplace operator with the three-dimensional one and use the Poisson equation (2.140),

$$\nabla_\perp \cdot \hat{\alpha} = 8\pi G \int \rho ds = 8\pi G \Sigma \,. \tag{2.161}$$

Σ is the *projected mass density* along the line of sight.

We can also write (2.153) as

$$\hat{\alpha} = 2\nabla_\perp \hat{\psi} \,, \quad \hat{\psi} = \int \phi ds \,, \tag{2.162}$$

where the integral (as in the previous equations) is always taken along the unperturbed path. Then we have

$$\Delta_\perp \hat{\psi} = 4\pi G\Sigma\,. \tag{2.163}$$

Using the Green's function $\mathcal{G}$ of the two-dimensional Laplacian,

$$\mathcal{G}(\boldsymbol{\xi}) = \frac{1}{2\pi}\ln|\boldsymbol{\xi}|\,, \tag{2.164}$$

where $\boldsymbol{\xi}$ is two dimensional coordinate vector in the transverse plane, we can write the projected potential $\hat{\psi}$ as

$$\hat{\psi}(\boldsymbol{\xi}) = 2G\int d^2\xi' \ln|\boldsymbol{\xi}-\boldsymbol{\xi}'|\Sigma(\boldsymbol{\xi}')\,, \tag{2.165}$$

and the deflection angle is then

$$\hat{\alpha} = 4G\int d^2\xi' \frac{\boldsymbol{\xi}-\boldsymbol{\xi}'}{|\boldsymbol{\xi}-\boldsymbol{\xi}'|^2}\Sigma(\boldsymbol{\xi}')\,. \tag{2.166}$$

For a point mass M located at the origin of the transversal plane, we have $\Sigma(\boldsymbol{\xi}) = M\delta(\boldsymbol{\xi})$, and we obtain

$$\hat{\alpha}(\boldsymbol{\xi}) = 4GM\frac{\hat{\boldsymbol{\xi}}}{|\boldsymbol{\xi}|}\,, \tag{2.167}$$

which is consistent with (2.159), with $|\boldsymbol{\xi}| = b$ and $\hat{\boldsymbol{\xi}}$ the unit vector pointing from the light ray position in the transverse plane ($\boldsymbol{\xi}$) towards the lens (at the origin).

Equation (2.166) is one of the key equations of gravitational lensing. We see that the deflection angle depends on the mass distribution (or more generally stress-energy distribution), which indicates that we should be able to constrain the distribution of mass between the source and the observer with the help of gravitational lensing. In addition, when relating the deflection angle to observables, we will also find geometric factors relating the distances between the source plane, the lens plane and the observer. Because of these factors, gravitational lensing is also able to constrain distances in the universe.

Chapter 3

The Schwarzschild solution, black holes and classical tests of GR

In the last chapter we have considered the weak-field limit of GR, which also provided us with an approximate metric for a point-like mass, Eq. (2.36). In this section we will limit ourselves to a static and spherically symmetric situation, and find the (unique) exact solution of Einstein's equation in vacuum, called the *Schwarzschild* metric (found by Karl Schwarzschild in 1916, only a few months after Einstein published his vacuum field equations). This solution is probably the second most important exact solution of GR, after the Minkowski metric.

We can then use the Schwarzschild solution to revisit predictions for the motion of bodies and light in the solar system, which provides us with precise predictions to test GR. We will also get insight into what happens if a body (like a large star) is no longer able to support itself against its self-gravity and undergoes complete gravitational collapse. We will see that GR predicts the existence of *black holes* that are described (in the simplest case) by the vacuum Schwarzschild solution everywhere.

But to start with, we have to define more precisely the meaning of 'static' and 'spherically symmetric'. To this end we will introduce some more mathematical machinery to deal with these notions in the first section.

3.1 Symmetries, the Lie derivative and Killing vectors

3.1.1 Pushforward and pullback

We return to the notion of a one-parameter family of diffeomorphisms $\phi_t : \mathcal{M} \to \mathcal{M}$ generated by a vector field v^μ, that we originally encountered at the end of Section 1.2.2[1]. The map $\phi : \mathcal{M} \to \mathcal{M}$ also defines a map (the *pushforward*) between tangent spaces, $\phi_* : V_p \to V_{\phi(p)}$: for a vector $\mathbf{v} \in V_p$ we define $\phi_* \mathbf{v} \in V_{\phi(p)}$ through

$$(\phi_* \mathbf{v})(f) = \mathbf{v}(f \circ \phi) \tag{3.1}$$

for all smooth functions $f : \mathcal{M} \to \mathbb{R}$. ϕ_* is a linear map, the matrix of its components (as it maps vectors on vectors) in the coordinate bases of a coordinate system $\{x^\nu\}$ at p and $\{y^\mu\}$ at $\phi(p)$ is the Jacobian matrix of the map ϕ between the coordinates, i.e. $(\phi_*)^\mu_\nu = \partial y^\mu / \partial x^\nu$.

The map ϕ also defines a map between the dual spaces (the *pullback*), $\phi^* : V^*_{\phi(p)} \to V^*_p$, through

$$(\phi^* \boldsymbol{\omega})_\mu v^\mu = \omega_\mu (\phi_* \mathbf{v})^\mu \tag{3.2}$$

[1] We could generalize the discussion to maps between two different manifolds, but we limit ourselves to just maps $\mathcal{M} \to \mathcal{M}$.

for all $\mathbf{v} \in V_p$. If a C^∞ map ϕ is a diffeomorphism then it is one-to-one, onto, and its inverse is C^∞. In this case it can be shown that $\phi^* = (\phi^{-1})_*$ so that it is enough to consider ϕ_* and $(\phi^{-1})_*$.

We can extend the action of a diffeomorphism ϕ to generally map tensors to tensors. If $T^{\mu_1\cdots\mu_k}_{\nu_1\cdots\nu_l}$ is a tensor of type (k,l) at p then we define the tensor $(\phi_* T)^{\mu_1\cdots\mu_k}_{\nu_1\cdots\nu_l}$ at $\phi(p)$ through

$$(\phi_*\mathbf{T})[\boldsymbol{\omega}_1, \cdots, \boldsymbol{\omega}_k, \mathbf{v}_1, \cdots, \mathbf{v}_l] = T[(\phi^*\boldsymbol{\omega}_1), \cdots, (\phi^*\boldsymbol{\omega}_k), ([\phi^{-1}]_*\mathbf{v}_1), \cdots, ([\phi^{-1}]_*\mathbf{v}_l)] \,. \tag{3.3}$$

If $\phi : \mathcal{M} \to \mathcal{M}$ is a diffeomorphism and $\mathbf{T}$ is a tensor field on $\mathcal{M}$ then we can compare $\phi_*\mathbf{T}$ with $\mathbf{T}$. If $\phi_*\mathbf{T} = \mathbf{T}$ even though we have 'moved' $\mathbf{T}$ with ϕ it has 'stayed the same'. In that case, ϕ is a *symmetry transformation* for the tensor field $\mathbf{T}$. A symmetry transformation for the metric, i.e. a diffeomorphism ϕ such that $(\phi_* g)_{\mu\nu} = g_{\mu\nu}$, is called an *isometry*.

3.1.2 The Lie derivative

Let $\mathcal{M}$ be a manifold and ϕ_t a one-parameter group of diffeomorphisms. As discussed in Section 1.2.2, ϕ_t will be generated by a vector field $\mathbf{v}$. We can then use ϕ_{t*} to carry along a smooth tensor field $\mathbf{T}$. Comparison of $\mathbf{T}$ and $\phi_*\mathbf{T}$ for $t \to 0$ defines a notion of derivative, called the *Lie derivative* $\mathcal{L}_\mathbf{v}$,

$$\mathcal{L}_\mathbf{v}\mathbf{T} = \lim_{t\to 0}\left(\frac{(\phi_*)_{-t}\mathbf{T} - \mathbf{T}}{t}\right) , \tag{3.4}$$

where all tensors are evaluated at the same point p. $\mathcal{L}_\mathbf{v}\mathbf{T} = 0$ everywhere if and only if for all t, ϕ_t is a symmetry transformation for $\mathbf{T}$.

It follows immediately from the definition that $\mathcal{L}_\mathbf{v}$ is a linear map from smooth tensor fields of type (k,l) to smooth tensor fields of type (k,l). Also, $\mathcal{L}_\mathbf{v}$ satisfies a Leibniz rule on outer products of tensors,

$$\mathcal{L}_\mathbf{v}(\mathbf{T}\otimes\mathbf{S}) = \mathcal{L}_\mathbf{v}\mathbf{T}\otimes\mathbf{S} + \mathbf{T}\otimes\mathcal{L}_\mathbf{v}\mathbf{S} \,. \tag{3.5}$$

As $\mathbf{v}$ is the tangent vector to the integral curves of ϕ_t, we have for functions $f : \mathcal{M} \to \mathbb{R}$ that

$$\mathcal{L}_\mathbf{v}(f) = \mathbf{v}(f) \,. \tag{3.6}$$

To simplify the analysis of the action of $\mathcal{L}_\mathbf{v}$ on tensors, it is useful to introduce 'adapted coordinates' on $\mathcal{M}$, where the parameter t along the integral curves of $\mathbf{v}$ is chosen as coordinate x^1, so that $\mathbf{v} = \boldsymbol{\partial}_1$ (which can always be done locally as long as $\mathbf{v} \neq 0$). The action of ϕ_{-t} then corresponds to the coordinate transformation $x^1 \to x^1 + t$, with $x^2, \ldots, x^n$ being held fixed. We then have $(\phi_*)^\mu_\nu = \delta^\mu_\nu$ and the coordinate basis components of $(\phi_*)_{-t}\mathbf{T}$ at the point p with coordinates $(x^1, \ldots, x^n)$ are

$$(\phi_*)_{-t}T^{\mu_1\cdots\mu_k}_{\nu_1\cdots\nu_l}(x^1, x^2, \ldots, x^n) = T^{\mu_1\cdots\mu_k}_{\nu_1\cdots\nu_l}(x^1 + t, x^2, \ldots, x^n) \,. \tag{3.7}$$

In a coordinate system adapted to $\mathbf{v}$ the Lie derivative of $\mathbf{T}$ in components is therefore just

$$(\mathcal{L}_\mathbf{v}\mathbf{T})^{\mu_1\cdots\mu_k}_{\nu_1\cdots\nu_l} = \frac{\partial T^{\mu_1\cdots\mu_k}_{\nu_1\cdots\nu_l}}{\partial x^1} \,. \tag{3.8}$$

One consequence is that ϕ_t will be a symmetry transformation of $\mathbf{T}$ if and only if the components $T^{\mu_1\cdots\mu_k}_{\nu_1\cdots\nu_l}$ in a coordinate system adapted to $\mathbf{v}$ are independent of the integral curve coordinate x^1.

To obtain a coordinate independent expression for the Lie derivative of a vector field $\mathbf{w}$ we note that in an adapted coordinate system we have

$$(\mathcal{L}_\mathbf{v}\mathbf{w})^\mu = \frac{\partial w^\mu}{\partial x^1} \,. \tag{3.9}$$

Since $\mathbf{v} = \boldsymbol{\partial}_1$ and $\mathbf{w} = \sum_\mu w^\mu \boldsymbol{\partial}_\mu$, the commutator of $\mathbf{v}$ and $\mathbf{w}$ is given by (1.27),

$$[\mathbf{v}, \mathbf{w}]^\mu = \sum_\nu \left(v^\nu \frac{\partial w^\mu}{\partial x^\nu} - w^\nu \frac{\partial v^\mu}{\partial x^\nu} \right) = \frac{\partial w^\mu}{\partial x^1} \,. \tag{3.10}$$

They are the same in this special coordinate system – but since both are defined in a coordinate independent way, they then have to be the same in any coordinate system, so that[2]

$$\mathcal{L}_\mathbf{v} \mathbf{w} = [\mathbf{v}, \mathbf{w}] \,. \tag{3.11}$$

We can then compute the action of the Lie derivative on other types of tensor fields from (3.6), (3.11) and the Leibniz rule. For example, for a dual vector field $\boldsymbol{\omega}$ we have

$$\mathcal{L}_\mathbf{v}(\omega_\mu w^\mu) = \mathbf{v}(\omega_\mu w^\mu) \,, \tag{3.12}$$

for an arbitrary vector field $\mathbf{w}$. But we also have, with the Leibniz rule,

$$\mathcal{L}_\mathbf{v}(\omega_\mu w^\mu) = \boldsymbol{\omega}(\mathcal{L}_\mathbf{v}\mathbf{w}) + (\mathcal{L}_\mathbf{v}\boldsymbol{\omega})(\mathbf{w}) = (\mathcal{L}_\mathbf{v}\boldsymbol{\omega})(\mathbf{w}) + \boldsymbol{\omega}([\mathbf{v}, \mathbf{w}]) \,. \tag{3.13}$$

By combining these expressions we obtain a formula for $\mathcal{L}_\mathbf{v}\boldsymbol{\omega}$. It is most convenient to express the result in terms of a derivative operator, for example the covariant derivative of Section 1.2.5 (but any derivative operator satisfying the necessary conditions works, also the usual derivative ∂). Using the Leibniz rule (1.46) and property (1.49) we can write

$$\mathbf{v}(\omega_\mu w^\mu) = v^\nu \nabla_\nu (\omega_\mu w^\mu) = v^\nu w^\mu \nabla_\nu \omega_\mu + v^\nu \omega_\mu \nabla_\nu w^\mu \,. \tag{3.14}$$

Using (1.64),

$$[\mathbf{v}, \mathbf{w}]^\nu = (v^\mu \nabla_\mu w^\nu - w^\mu \nabla_\mu v^\nu) = (\mathcal{L}_\mathbf{v}\mathbf{w})^\nu \,, \tag{3.15}$$

we can write

$$v^\nu w^\mu \nabla_\nu \omega_\mu + v^\nu \omega_\mu \nabla_\nu w^\mu = w^\mu (\mathcal{L}_\mathbf{v}\omega)_\mu + \omega_\mu v^\nu \nabla_\nu w^\mu - \omega_\mu w^\nu \nabla_\nu v^\mu \,, \tag{3.16}$$

or

$$(\mathcal{L}_\mathbf{v}\omega)_\mu = v^\nu \nabla_\nu \omega_\mu + \omega_\nu \nabla_\mu v^\nu \,. \tag{3.17}$$

The general expression is (again expressed using ∇, but we could have used ∂),

$$(\mathcal{L}_\mathbf{v}\mathbf{T})^{\mu_1 \cdots \mu_k}_{\nu_1 \cdots \nu_l} = v^\rho \nabla_\rho T^{\mu_1 \cdots \mu_k}_{\nu_1 \cdots \nu_l} - \sum_{i=1}^k T^{\mu_1 \cdots \rho \cdots \mu_k}_{\nu_1 \cdots \nu_l} \nabla_\rho v^{\mu_i} + \sum_{j=1}^l T^{\mu_1 \cdots \mu_k}_{\nu_1 \cdots \rho \cdots \nu_l} \nabla_{\nu_j} v^\rho \,. \tag{3.18}$$

Specifically for the metric tensor we have

$$(\mathcal{L}_\mathbf{v} g)_{\mu\nu} = v^\rho \nabla_\rho g_{\mu\nu} + g_{\rho\nu} \nabla_\mu v^\rho + g_{\mu\rho} \nabla_\nu v^\rho = \nabla_\mu v_\nu + \nabla_\nu v_\mu \tag{3.19}$$

where the last equality holds if ∇ is the derivative associated with the metric $g_{\mu\nu}$.

[2] It is then also easy to see with (1.64) that for a torsion-free connection $\mathcal{L}_\mathbf{v}\mathbf{w} = \nabla_\mathbf{v}\mathbf{w} - \nabla_\mathbf{w}\mathbf{v}$.

3.1.3 Killing vectors

If $\phi_t : \mathcal{M} \to \mathcal{M}$ is a one-parameter group of isometries then the vector field ξ^μ which generates ϕ_t is called a *Killing vector field*. As mentioned above, the necessary and sufficient condition for ϕ_t to be a group of isometries is $\mathcal{L}_{\boldsymbol{\xi}} g_{\mu\nu} = 0$. With (3.19) this implies that the necessary and sufficient condition for ξ^μ to be a Killing vector field is for it to satisfy *Killing's equation*,

$$\nabla_\mu \xi_\nu + \nabla_\nu \xi_\mu = 0\,, \tag{3.20}$$

where ∇ is the derivative operator associated with the metric $g_{\mu\nu}$.

A useful property of Killing vector fields is the following: Let $\boldsymbol{\xi}$ be a Killing vector field and γ a geodesic with tangent $\mathbf{u}$. Then $\xi_\mu u^\mu$ is constant along γ. This can be seen by direct calculation:

$$u^\nu \nabla_\nu (\xi_\mu u^\mu) = u^\nu u^\mu \nabla_\nu \xi_\mu + \xi_\mu u^\nu \nabla_\nu u^\mu = 0\,, \tag{3.21}$$

since the first term vanishes due to Killing's equation (3.20) and second term vanishes from the geodesic equation.

The interpretation of this property is that a one-parameter family of symmetries gives rise to a conserved quantity for particles and light rays since both follow (time-like and null) geodesics. This can be useful in many situations where symmetries are present.

Example: Killing vectors of Minkowski space-time

The Minkowski metric is usually given by

$$ds^2 = \eta_{\mu\nu} dx^\mu dx^\nu \tag{3.22}$$

for $x^\mu = \{t, x, y, z\}$. As discussed in Chapter 1, Minkowski space-time is flat, $\Gamma^\lambda_{\mu\nu} = 0$ and thus $\nabla_\mu = \partial_\mu$. Geodesics satisfy $u^\mu \partial_\mu u^\nu = 0$, or in coordinates

$$\frac{d^2 x^\mu(\lambda)}{d\lambda^2} = 0\,, \tag{3.23}$$

so geodesics in Minkowski are straight lines $x^\mu(\lambda) = x_0^\mu + \lambda u^\mu$ where u^μ is a constant, normalised vector (field). The equation for a Killing vector field $\boldsymbol{\xi}$ is simply

$$\partial_\mu \xi_\nu + \partial_\nu \xi_\mu = 0\,. \tag{3.24}$$

We also see that the metric is independent of all coordinates and thus these are adapted coordinates for isometries in the coordinate directions, and so

$$P_0 = \boldsymbol{\partial}_t\,, \quad P_1 = \boldsymbol{\partial}_x\,, \quad P_2 = \boldsymbol{\partial}_y\,, \quad P_3 = \boldsymbol{\partial}_z \tag{3.25}$$

are Killing vector fields for the canonical coordinates of Minkowski space (corresponding to $\xi^\mu = (1, 0, 0, 0)$, etc). They satisfy Killing's equation trivially. The associated conserved quantities are $\xi_\mu u^\mu$, which simply projects out (up to a proportionality constant, for a massive particle $p^\mu = m u^\mu$) the energy and momentum of the four velocity u^μ in the coordinate rest-frame, quantities that are indeed conserved in a Minkowski space-time.

But we could also have used spherical coordinates,

$$ds^2 = -dt^2 + dr^2 + r^2 \left(d\theta^2 + \sin^2\theta d\varphi^2\right)\,. \tag{3.26}$$

The metric components do not depend on the angular variables, and the spatial part of Minkowski is spherically symmetric. This provides three additional Killing vector fields that were not trivially apparent in the canonical coordinates above. They are given by the generators of the (spatial) rotations, explicitly they are

$$J_1 = -z\boldsymbol{\partial}_y + y\boldsymbol{\partial}_z\,, \quad J_2 = -x\boldsymbol{\partial}_z + z\boldsymbol{\partial}_x\,, \quad J_3 = -y\boldsymbol{\partial}_x + x\boldsymbol{\partial}_y\,. \tag{3.27}$$

Again it is simple to verify Killing's equation. The associated conserved quantity is the angular momentum $\mathbf{L}$.

But as we briefly discussed in Section 1.3.1, the Minkowski metric has a larger invariance group, namely the Lorentz group, which includes also boosts, generated by the vector fields

$$K_1 = x\boldsymbol{\partial}_t + t\boldsymbol{\partial}_x\,, \quad K_2 = y\boldsymbol{\partial}_t + t\boldsymbol{\partial}_y\,, \quad K_3 = z\boldsymbol{\partial}_t + t\boldsymbol{\partial}_z\,. \tag{3.28}$$

The associated conserved quantities, when combined with the usual angular momentum, can be seen as a kind of 'relativistic angular momentum' as the boosts are effectively 'rotations' involving a time and a space direction (similar to combining energy and momentum to the relativistic 4-momentum). Since time is necessarily involved in boosts, the resulting conserved quantity is 'time dependent' which is a bit confusing, and is probably why it is rarely used or discussed in mechanics texts.

Overall we thus have 10 Killing vectors for the Minkowski space-time. This is actually the maximal number of Killing vectors for a four-dimensional space-time (it can be shown that in n dimensions the maximum is $n(n+1)/2$), and Minkowski is for this reason called a maximally symmetric space-time.

3.2 The Schwarzschild metric

We would like to find all vacuum solutions of GR that are static and spherically symmetric. The first step is to define what we mean with static and spherically symmetric, and to find the most general metric that satisfies these properties. In a second step we then need to find a solution of the vacuum Einstein equations $R_{\mu\nu} = 0$ for such a metric.

A space-time is called *stationary* if it possesses a one-parameter group of isometries ϕ_t whose orbits are time-like curves. This group of isometries expresses the 'time translation symmetry' of the space-time. From Section 3.1.3 we know that this is equivalent to the existence of a time-like Killing vector field $\mathbf{T}$.

A space-time is called *static* if it is stationary and in addition there exists a (space-like) hypersurface Σ which is orthogonal to the orbits of the isometry. In this case we can introduce specific coordinates for the static space-time. If $\mathbf{T} \neq 0$ everywhere on Σ then in a neighbourhood of Σ every point will lie on unique orbit generated by $\mathbf{T}$, passing through Σ. We then choose arbitrary coordinates $\{x^i\}$ on Σ and label each point p in the neighbourhood of Σ by the parameter t of the orbit which starts from Σ and ends on p, and the coordinates $\{x^1, x^2, x^3\}$ of the orbit at Σ. From the discussion around Eq. (3.8) and Section 3.1.3, since this coordinate system uses the Killing parameter t as one of the coordinates, the metric components in this coordinate basis will be independent of t. We define the surface Σ_t as the set of points with 'time coordinate' t, i.e. the image of Σ under the isometry ϕ_t, so that each Σ_t is also orthogonal to $\mathbf{T}$. Therefore, in these coordinates the metric takes the form

$$ds^2 = -V^2(x^1, x^2, x^3)dt^2 + \sum_{i,j=1}^{3} h_{ij}(x^1, x^2, x^3)dx^i dx^j\,, \tag{3.29}$$

with $V^2 = -T_\mu T^\mu$. The absence of $dtdx^i$ cross terms expresses the orthogonality of $\mathbf{T}$ with Σ. Conversely, a stationary but non-static metric necessarily has $dtdx^i$ cross terms in any coordinate system that uses the Killing parameter as one of the coordinates.

From the explicit expression for a static metric given above, we see that the diffeomorphism $t \to -t$, i.e. the map that takes each point on Σ_t to the point with the same spatial coordinates on Σ_{-t}, is an isometry. Thus, a static space-time also has a time reflection symmetry in addition to a time translation symmetry. An example where this is not the case is the presence of rotating motion. Although e.g. a rotating fluid ball can have a time-independent matter and velocity distribution, and thus be time translation invariant, it does not possess time reflection symmetry as in this case the direction of rotation changes.

A space-time is called *spherically symmetric* if its isometry group contains a subgroup that is isomorphic to the rotation group SO(3), and the orbits of that subgroup are two-dimensional spheres. Physically we can interpret the action of the isometry group as rotations, and spherically symmetric space-times as space-times with a metric that remains invariant under rotations. The space-time metric induces a metric on each of these orbit 2-spheres. Because of the rotational symmetry, the induced metric must be a multiple of the metric of a unit 2-sphere, and therefore characterised by the total area A of each 2-sphere. It is then convenient for spherically symmetric space-times to introduce the function r through

$$r \equiv \sqrt{\frac{A}{4\pi}}\,. \tag{3.30}$$

Using spherical coordinates (θ, ϕ), the metric on each orbit 2-sphere takes the form

$$ds_2^2 = r^2(d\theta^2 + \sin^2\theta d\phi^2) = r^2 d\Omega^2\,. \tag{3.31}$$

We should note that r does not need to be a radius from a center – indeed, the space-time manifold may not have a center, e.g. if it is $\mathbb{R} \times S^2$ rather than $\mathbb{R}^3$. However we will anyway call r the 'radial coordinate' of the sphere.

If a space-time is both static and spherically symmetric, and if the static Killing vector field $\mathbf{T}$ is unique, then $\mathbf{T}$ is orthogonal to the two spheres: a unique $\mathbf{T}$ has to be invariant under the rotational isometries, which requires that its projection onto an orbit 2-sphere has to vanish since otherwise it cannot be invariant under all rotations. The orbit 2-spheres therefore lie within the hypersurfaces Σ_t that are orthogonal to $\mathbf{T}$. We can choose convenient coordinates as follows: We select one of the spheres on $\Sigma = \Sigma_{t=0}$ and choose spherical coordinates (θ, ϕ) on this sphere. We then transport those spherical coordinates to the other spheres of Σ, using geodesics orthogonal to the 2-sphere. We can now use coordinates (r, θ, ϕ) on Σ with r defined in Eq. (3.30). Finally we can transfer those coordinates to Σ_t using $\mathbf{T}$, and choose (t, r, θ, ϕ) as coordinates on the space-time. In these coordinates, the metric of an arbitrary static, spherically symmetric space-time takes the simple form

$$ds^2 = -f(r)dt^2 + h(r)dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2) = -e^{2\alpha(r)}dt^2 + e^{2\beta(r)}dr^2 + r^2 d\Omega^2\,. \tag{3.32}$$

Writing the free functions $f(r)$ and $h(r)$ as exponentials is convenient because they are automatically positive and thus prevent accidentally changing the signature of the metric. We should also note that the coordinate system breaks down in some places, for example at the north and south poles of the 2-spheres (as usual) but also in places where $\mathbf{T} = 0$ or $\nabla_\mu r = 0$ (or more generally where T^μ and $\nabla^\mu r$ become collinear). We will revisit this later.

We now have reduced the problem of finding 10 unknown functions $g_{\mu\nu}$ of four variables x^μ to the much easier problem of determining two functions α and β of only one variable, r. To do this, we have to compute the Ricci tensor $R_{\mu\nu}$ of the metric (3.32) and to solve $R_{\mu\nu} = 0$ for $\alpha(r)$ and $\beta(r)$. There are various ways to do this, we use the most basic approach, following [4], by just explicitly computing the Christoffel symbols Γ and the components of the Ricci tensor.

The non-zero Christoffel symbols (up to those related by symmetries) are found to be

$$\begin{aligned} &\Gamma^t_{tr} = \partial_r\alpha && \Gamma^r_{tt} = e^{2(\alpha-\beta)}\partial_r\alpha && \Gamma^r_{rr} = \partial_r\beta \\ &\Gamma^\theta_{r\theta} = 1/r && \Gamma^r_{\theta\theta} = -re^{-2\beta} && \Gamma^\phi_{r\phi} = 1/r \\ &\Gamma^r_{\phi\phi} = -re^{-2\beta}\sin^2\theta && \Gamma^\theta_{\phi\phi} = -\sin\theta\cos\theta && \Gamma^\phi_{\theta\phi} = \frac{\cos\theta}{\sin\theta}\,. \end{aligned} \tag{3.33}$$

The non-vanishing components of the Riemann tensor (again up those determined by symmetries) are then

$$\begin{aligned} &R^t_{rtr} = \partial_r\alpha\partial_r\beta - \partial_r^2\alpha - (\partial_r\alpha)^2 && R^t_{\theta t\theta} = -re^{-2\beta}\partial_r\alpha \\ &R^t_{\phi t\phi} = -re^{-2\beta}\sin^2\theta\partial_r\alpha && R^r_{\theta r\theta} = re^{-2\beta}\partial_r\beta \\ &R^r_{\phi r\phi} = re^{-2\beta}\sin^2\theta\partial_r\beta && R^\theta_{\phi\theta\phi} = \left(1-e^{-2\beta}\right)\sin^2\theta\,. \end{aligned} \tag{3.34}$$

The Ricci tensor, computed as the contraction of the Riemann tensor, is

$$R_{tt} = e^{2(\alpha-\beta)}\left[\partial_r^2\alpha + (\partial_r\alpha)^2 - \partial_r\alpha\partial_r\beta + \frac{2}{r}\partial_r\alpha\right] \tag{3.35}$$

$$R_{rr} = -\partial_r^2\alpha - (\partial_r\alpha)^2 + \partial_r\alpha\partial_r\beta + \frac{2}{r}\partial_r\beta \tag{3.36}$$

$$R_{\theta\theta} = e^{-2\beta}\left[r(\partial_r\beta - \partial_r\alpha) - 1\right] + 1 \tag{3.37}$$

$$R_{\phi\phi} = \sin^2\theta R_{\theta\theta}\,, \tag{3.38}$$

All components of the Ricci tensor need to be equal to zero. Since R_{tt} and R_{rr} vanish independently, we can write

$$0 = e^{2(\beta-\alpha)}R_{tt} + R_{rr} = \frac{2}{r}(\partial_r\alpha + \partial_r\beta)\,, \tag{3.39}$$

which implies that $\alpha = -\beta + c$. The constant c can be set to zero by rescaling the time coordinate as $t \to e^{-c}t$, so that

$$\alpha = -\beta\,. \tag{3.40}$$

Using this in the $R_{\theta\theta} = 0$ equations gives

$$e^{2\alpha}(2r\partial_r\alpha + 1) = 1\,. \tag{3.41}$$

This can be written as

$$\partial_r\left(re^{2\alpha}\right) = 1\,, \tag{3.42}$$

which is solved by

$$e^{2\alpha} = 1 - \frac{R_S}{r}\,. \tag{3.43}$$

Here R_S is an unknown constant. Our metric is then given by

$$ds^2 = -\left(1-\frac{R_S}{r}\right)dt^2 + \left(1-\frac{R_S}{r}\right)^{-1}dr^2 + r^2d\Omega^2\,. \tag{3.44}$$

This metric solves all equations $R_{\mu\nu} = 0$ above, for all values of R_S.

The constant R_S is called the *Schwarzschild radius*. To understand its significance better, we can compare the *Schwarzschild metric* (3.44) to the metric of a point-mass in the weak field limit, Eq. (2.36). There we found that

$$g_{tt} = -\left(1 - 2\frac{GM}{r}\right)\,. \tag{3.45}$$

Since they are of the same form, we can identify

$$R_S = 2GM\,. \tag{3.46}$$

We also see that in the weak-field limit $r \gg 2GM$ of the Schwarzschild metric we do indeed recover at first order the metric (2.36). The above identification is thus consistent. The Schwarzschild metric is usually written in terms of M,

$$ds^2 = -\left(1 - \frac{2GM}{r}\right) dt^2 + \left(1 - \frac{2GM}{r}\right)^{-1} dr^2 + r^2 d\Omega^2\,. \tag{3.47}$$

M here is a parameter that can be interpreted with the 'Newtonian mass' that we would infer from the trajectory of inertial observers at large distances from the gravitating source. For $M \to 0$ we recover the Minkowski space-time. For $r \to \infty$ the metric also approaches Minkowski, a property called *asymptotic flatness*.

Looking at (3.44) we can't help noticing that the metric diverges when $r = 0$ and when $r = R_S$. This does not necessarily mean that there is a problem at these locations, as this may be just an artefact of our choice of coordinates (a *coordinate singularity*). An example is the origin of polar coordinates in the plane, with the metric $ds^2 = dr^2 + r^2 d\theta^2$. There the component $g^{\theta\theta} = 1/r^2$ of the inverse metric diverges, even though we know that the plane as a manifold is perfectly regular and that all points (including the arbitrary choice of an origin) are the same.

Giving a general diagnostic for a physical singularity in GR is difficult. A singularity in a physically meaningful quantity would certainly signal the presence of a problem. An example would be a scalar constructed from the Riemann curvature tensor, as the components of a tensor are coordinate dependent, while scalars are not. For a vacuum metric the Ricci scalar vanishes by construction, $R = 0$. We can however construct other scalars, for example the *Kretschmann scalar* $K = R_{\mu\nu\lambda\rho}R^{\mu\nu\lambda\rho}$, but there are many other possibilities. If any such scalar diverges at a point p of the manifold then we consider that point as being singular (although in general we should also check that it can be reached with a geodesic of finite length).

For the Schwarzschild metric (3.47) we find by direct computation that

$$K = R_{\mu\nu\lambda\rho}R^{\mu\nu\lambda\rho} = \frac{48G^2M^2}{r^6}\,. \tag{3.48}$$

This indicates that $r = 0$ is a real singularity. At $r = 2GM$ on the other hand, this and other such scalars remain finite. We will see later that $r = 2GM$ is indeed a coordinate singularity and that we can find coordinates where nothing blows up there. The surface $r = R_S$ is nonetheless interesting as it is the event horizon of a black hole, as we will also discuss later.

As a final remark, we could also have studied just a spherically symmetric vacuum solution of Einstein's field equations, without requiring it to be static. We would have found the Schwarzschild solution as well, a fact known as *Birkhoff's theorem* (1923). More details and the calculation can be found for example in [2, 4]. Therefore all spherically symmetric vacuum solutions of GR are static and have a time-like killing vector, and actually there is a unique family of spherically symmetric vacuum solutions given by the metric (3.47). This is once again closely analogous to electrodynamics where the Coulomb solution is the only spherically symmetric solution of Maxwell's equation. In both cases it can be interpreted as saying that there is no (gravitational or electromagnetic) monopole (spherically symmetric) radiation – electrodynamic radiation couples to a dipolar source, while, as we have seen, gravitational waves couple to the quadrupole of the source. Thanks to Birkhoff's theorem, the Schwarzschild metric is generally the exterior field also for non-static mass distributions. For this reason an exactly spherically symmetric stellar explosion or collapse would not generate any gravitational waves.

3.3 Geodesics of the Schwarzschild metric

The Schwarzschild solution describes for example the gravitational field of the sun very well. It should thus allow us to predict the motion of the planets, and to find out why the perihelion of Mercury precesses more than what is expected from the influence of the other planets. Mercury's anomalous rate of precession was known since 1859, from an analysis by Urbain Le Verrier. In this section we will study the general properties of the geodesics in a Schwarzschild metric and then return to the Mercury perihelion question in the next section.

The geodesic equation can be found by simply using the Christoffel symbols (3.33) for the functions α and β of the Schwarzschild metric, given by (3.43) and (3.40). We find the four coupled differential equations

$$\frac{d^2t}{d\lambda^2} + \frac{2GM}{r(r-2GM)}\frac{dr}{d\lambda}\frac{dt}{d\lambda} = 0\,, \tag{3.49}$$

$$\begin{aligned} \frac{d^2r}{d\lambda^2} + \frac{GM}{r^3}(r-2GM)\left(\frac{dt}{d\lambda}\right)^2 - \frac{GM}{r(r-2GM)}\left(\frac{dr}{d\lambda}\right)^2 & \\ -(r-2GM)\left[\left(\frac{d\theta}{d\lambda}\right)^2 + \sin^2\theta\left(\frac{d\phi}{d\lambda}\right)^2\right]^2 &= 0\,, \end{aligned} \tag{3.50}$$

$$\frac{d^2\theta}{d\lambda^2} + \frac{2}{r}\frac{d\theta}{d\lambda}\frac{dr}{d\lambda} - \sin\theta\cos\theta\left(\frac{d\phi}{d\lambda}\right)^2 = 0\,, \tag{3.51}$$

$$\frac{d^2\phi}{d\lambda^2} + \frac{2}{r}\frac{d\phi}{d\lambda}\frac{dr}{d\lambda} + 2\frac{\cos\theta}{\sin\theta}\frac{d\theta}{d\lambda}\frac{d\phi}{d\lambda} = 0\,. \tag{3.52}$$

This system of equations could of course be solved numerically, but that would not provide much insight. Luckily, as mentioned in Section 3.1.3, we can use the symmetries of the Schwarzschild metric, and the associated conserved quantities, to radically simplify the problem.

Each killing vector ξ^μ of the Schwarzschild metric will lead to a constant of motion for a free particle through (3.21), or equivalently

$$\xi_\mu \frac{dx^\mu}{d\lambda} = \text{const. (along geodesic)} \tag{3.53}$$

We note that in addition to these constants of motion, a geodesic also preserves the normalisation

$$\epsilon = -g_{\mu\nu}\frac{dx^\mu}{d\lambda}\frac{dx^\nu}{d\lambda}\,. \tag{3.54}$$

For massive particles following time-like geodesics we usually choose $\epsilon = 1$, implying $\lambda = \tau$. For massless particles traveling on null geodesics the normalisation is $\epsilon = 0$ which does not fix λ, instead we can choose $p^\mu = u^\mu = dx^\mu/d\lambda$, i.e. four-momentum and four-velocity are equal.

Intuitively we expect that the invariance under time translations should lead to energy (E) conservation, and the invariance under spatial rotation to conservation of angular momentum. The latter corresponds actually to three conservation laws, one for each component of the angular momentum vector $\mathbf{L}$, which we can write in terms of its magnitude L and two directions.

The conservation of the direction of angular momentum means that the particle moves in a plane. We can always choose our coordinate system so that this is the equatorial plane,

$$\theta = \frac{\pi}{2}\,. \tag{3.55}$$

We can actually see immediately from Eq. (3.51) that $d\theta/d\lambda = 0$ is a solution in this case, since in the equatorial plane $\cos\theta = 0$. Therefore a trajectory that is initially in the equatorial plane will remain in the equatorial plane.

The time-like Killing vector (field) is

$$K^\mu = (\boldsymbol{\partial}_t)^\mu = (1,0,0,0)\,, \tag{3.56}$$

and the Killing vector corresponding to the conservation of L is

$$R^\mu = (\boldsymbol{\partial}_\phi)^\mu = (0,0,0,1)\,. \tag{3.57}$$

Lowering the indices, we have

$$K_\mu = \left(-\left(1-\frac{2GM}{r}\right),0,0,0\right)\,, \quad R_\mu = \left(0,0,0,r^2\sin^2\theta\right)\,. \tag{3.58}$$

It is straightforward to check that both vector fields satisfy Kiling's equation (3.20). From (3.53), and using that $\sin\theta = 1$ for motion in the equatorial plane, we then find the conserved quantities

$$E = -K_\mu\frac{dx^\mu}{d\lambda} = \left(1-\frac{2GM}{r}\right)\frac{dt}{d\lambda}\,, \tag{3.59}$$

$$L = R_\mu\frac{dx^\mu}{d\lambda} = r^2\frac{d\phi}{d\lambda}\,. \tag{3.60}$$

Returning to the normalisation of the geodesics (3.54), we have in more detail

$$-\left(1-\frac{2GM}{r}\right)\left(\frac{dt}{d\lambda}\right)^2 + \left(1-\frac{2GM}{r}\right)^{-1}\left(\frac{dr}{d\lambda}\right)^2 + r^2\left(\frac{d\phi}{d\lambda}\right)^2 = -\epsilon\,. \tag{3.61}$$

Multiplying this expression by $(1-2GM/r)$ and using our expressions for E and L we find

$$-E^2 + \left(\frac{dr}{d\lambda}\right)^2 + \left(1-\frac{2GM}{r}\right)\left(\frac{L^2}{r^2}+\epsilon\right) = 0\,. \tag{3.62}$$

We can rewrite this equation suggestively as

$$\frac{1}{2}\left(\frac{dr}{d\lambda}\right)^2 + V(r) = \mathcal{E} \tag{3.63}$$

for the effective potential and 'effective energy'

$$V(r) = \frac{1}{2}\left(\epsilon+\frac{L^2}{r^2}\right)\left(1-\frac{2GM}{r}\right) = \frac{1}{2}\epsilon - \epsilon\frac{GM}{r} + \frac{L^2}{2r^2} - \frac{GML^2}{r^3}\,, \tag{3.64}$$

$$\mathcal{E} = \frac{1}{2}E^2\,. \tag{3.65}$$

Equation (3.63) is just the classical energy equation for a particle of mass $m = 1$ and energy $\mathcal{E}$ moving in one dimension in a potential given by $V(r)$. Of course we should not forget that this only gives us the radial motion, in reality the trajectories (e.g. the approximate ellipses of planets around the sun) also depend on $t(\lambda)$ and $\phi(\lambda)$ in addition to $r(\lambda)$. These two functions can be found e.g. from (3.59) and (3.60).

In the effective potential (3.64), the first term of the last equation is just a constant. The second term corresponds precisely to the Newtonian gravitational potential, and the third term to a contribution from

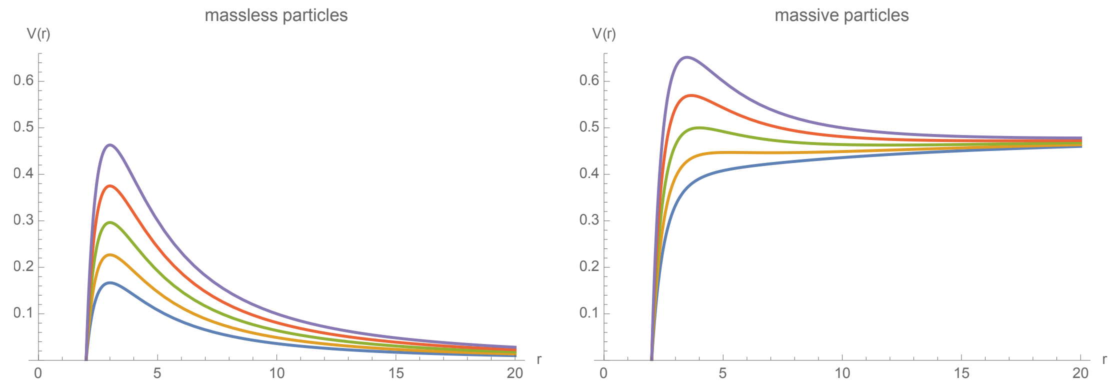

Figure 3.1: The effective GR potential $V(r)$, Eq. (3.64), for massless particles ($\epsilon = 0$) on the left and for massive particles ($\epsilon = 1$) on the right. In both cases the curves are for an angular momentum per unit mass $L = 3, 3.5, 4, 4.5, 5$ (from bottom to top), and we set $GM = 1$. Any particle on a geodesic that falls inside the innermost circular orbit at $r = 3GM$ necessarily continues to $r = 0$.

the angular momentum that is also present in Newtonian gravity. All of these terms we would also have found in a Newtonian analysis, the exception is the last term, a ‘pure’ GR contribution that will always dominate for small radii.

We show the effective potential for some values of the angular momentum per unit mass L in Fig. 3.1. The overall behaviour of $r(\lambda)$ can be deduced by comparing $\mathcal{E}$ to to $V(r)$. For a particle coming in from infinity with an energy $\mathcal{E}$, for a potential that grows as r decreases like the top-most curve of the right-hand plot (for $L = 5$), the particle will reach a minimal r at a turning point where $\mathcal{E} = V(r)$ and $dr/d\lambda = 0$ (only ‘potential energy’, no ’kinetic energy’). It will then move away again. If the energy is too high, or if the potential is not actually increasing, then the particle will just keep going toward $r = 0$.

Another possibility is a bound orbit, with circular orbits being a special case. On a circular orbit, $r = r_c$ remains constant. This can happen (depending on the initial conditions) at the places where the potential is flat. Differentiating (3.64) we see that circular orbits are possible when

$$\epsilon GM r_c^2 - L^2 r_c + 3GML^2\gamma = 0 , \tag{3.66}$$

where $\gamma = 0$ in Newtonian physics and $\gamma = 1$ in GR. Circular orbits will be stable if they lie at the minimum of a potential, and unstable if they occur at a maximum of a potential. Bound orbits that are not circular will oscillate between two radii with identical $V(r)$, around the radius of a stable circular orbit.

In Newtonian gravity, circular orbits exist for

$$r_c = \frac{L^2}{\epsilon GM} . \tag{3.67}$$

There are no stable circular orbits for massless particles ($\epsilon = 0$), which is not surprising. Indeed there are no bound orbits. In Newtonian gravity massless particles move on straight lines[3] as the Newtonian gravitational force GMm/r^2 vanishes for a particle with $m = 0$. The apparently repulsive potential as a function of L is just due to particles coming closer to the point mass the closer their initial velocity vector is pointing toward it. Massive particles can follow circular orbits for r_c given by (3.67) with $\epsilon = 1$. We know that in general particles in Newtonian gravity follow conic sections, i.e. their trajectories follow closed ellipses, parabolas or hyperbolas, in addition to circles.

[3]This can change depending on assumptions made, in some approximations light rays are also deflected by a Newtonian gravitational potential, but only by half the amount as light in GR. Here the Newtonian gravitational potential is ‘turned off’ if $\epsilon = 0$.

For $r \to \infty$ the effective potential in GR and in Newtonian gravity is the same, but at small r it is radically different. In the Newtonian case, without the last term of expression (3.64), the potential grows without limit as $r \to 0$ so that all particles with $L > 0$ will turn around at some point. In other words, in Newtonian physics a particle can never 'fall into' a point mass unless it is directly aimed at it. In GR on the other hand the term proportional to $-1/r^3$ will eventually dominate. At $r = 2GM$ the potential is always zero.

As Fig. 3.1 shows, there is always a barrier for massless particles. It is however of finite height so that sufficiently energetic particles can overcome it and fall toward $r \to 0$. The top of the barrier corresponds to an unstable circular orbit, at

$$r_c = 3GM \tag{3.68}$$

independently of L.

For massive particles, the circular orbits are at

$$r_c = \frac{L^2 \pm \sqrt{L^4 - 12G^2M^2L^2}}{2GM} . \tag{3.69}$$

For large L (to leading order in GM/L) there are two circular orbits, a stable orbit at $r_c = L^2/(GM)$ and an unstable orbit at $r_c = 3GM$. Thus, as L increases, the stable orbit moves further away from the center, while the unstable orbit approaches the same r_c as for the massless particles. As L decreases, the two orbits approach each other and they coincide when the square root in (3.69) vanishes, where

$$L = 2\sqrt{3}GM\,, \quad r_c = 6GM\,. \tag{3.70}$$

For smaller L there are no circular orbits as the potential is never flat but instead always has a positive slope. Therefore $6GM$ is the smallest stable circular orbit in the Schwarzschild metric (often called 'innermost stable circular orbit' or ISCO). For $L > 2\sqrt{3}GM$ there are also unbound orbits that come in from infinity and turn around, and bound but non-circular orbits that oscillate around the stable circular orbit. We will see in the next section that in GR these orbits are not exact conical sections.

Finally we should remember that these orbits are all geodesics. Although all geodesics will fall toward $r \to 0$ if they enter $r < 3GM$, it is not impossible to reach the region $r < 3GM$ (but $r > 2GM$ as we will see) and to escape to infinity by accelerating. We should also not forget that the Schwarzschild metric is the exterior field for both spherically symmetric stars and black holes. So it would not make a difference (in terms of the geodesic) whether the Earth is orbiting the Sun or an equally massive black hole. Dramatic things only start to happen as we approach the Schwarzschild radius, but for stars $R_S \ll R_*$, and the internal structure of a star is entirely regular, there is no singularity or horizon hidden there.

3.4 Classical tests of GR

3.4.1 Precession of the perihelion of a planet

Because of the last, 'non-Newtonian' term in the effective potential (3.64), 'elliptical' orbits in GR are not closed. In the solar system they can be well described by precessing ellipses, describing a flower-like pattern shown in Fig. 3.2.

To study the precession, we would like to find an equation for $r(\phi)$. To this end, we multiply the radial equation of motion (3.63) with $2(d\phi/d\lambda)^{-2} = 2r^4/L^2$, and find for $\epsilon = 1$

$$\left(\frac{dr}{d\phi}\right)^2 + \frac{1}{L^2}r^4 - \frac{2GM}{L^2}r^3 + r^2 - 2GMr = \frac{2\mathcal{E}}{L^2}r^4\,. \tag{3.71}$$

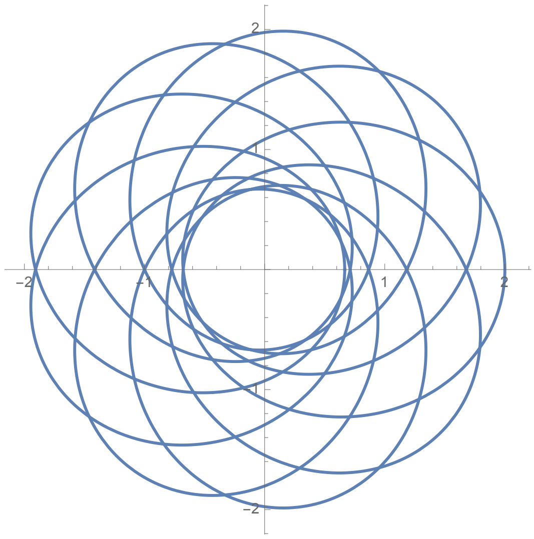

Figure 3.2: A (somewhat exaggerated) example of the precession of elliptical orbits in GR, using the approximation (3.81) for $e = 0.5$ and $\alpha = 0.1$. In the solar system the effect is much smaller, but it can become large near very compact objects.

We introduce a new variable which replaces r (analogously to the standard approach for solving the Kepler problem)

$$x = \frac{L^2}{GMr}\,. \tag{3.72}$$

From (3.67) we see that for a circular Newtonian orbit $x = 1$. In terms of x our equation becomes

$$\left(\frac{dx}{d\phi}\right)^2 + \frac{L^2}{G^2M^2} - 2x + x^2 - \frac{2G^2M^2}{L^2}x^3 = \frac{2\mathcal{E}L^2}{G^2M^2}\,. \tag{3.73}$$

By differentiating this equation with respect to ϕ we remove the constant terms and obtain a second order equation of motion,

$$\frac{d^2x}{d\phi^2} - 1 + x = \frac{3G^2M^2}{L^2}x^2\,. \tag{3.74}$$

The term on the right-hand side is the new contribution in GR. Without this term there would be a simple solution. But the right-hand term will generally be small in the solar system, so that we can split the total $x(\phi)$ into a Newtonian part and a perturbation,

$$x = x_0 + x_1\,. \tag{3.75}$$

The Newtonian part then is the solution to the equation

$$\frac{d^2x_0}{d\phi^2} - 1 + x_0 = 0\,, \tag{3.76}$$

which is

$$x_0 = 1 + e\cos\phi\,. \tag{3.77}$$

The result describes an ellipse with eccentricity e (or a hyperbola or parabola if $e \geq 1$), and the two initial conditions specify the orientation of the conic section as well as e. For an ellipse with semi-major axis a and semi-minor axis b, the eccentricity is given by $e^2 = 1 - b^2/a^2$.

The first-order equation is

$$\frac{d^2x_1}{d\phi^2} + x_1 = \frac{3G^2M^2}{L^2}x_0^2\,. \tag{3.78}$$

Inserting the Newtonian ellipse (3.77) gives

$$\frac{d^2x_1}{d\phi^2} + x_1 = \frac{3G^2M^2}{L^2}(1+e\cos\phi)^2 = \frac{3G^2M^2}{L^2}\left[\left(1+\frac{1}{2}e^2\right) + 2e\cos\phi + \frac{1}{2}e^2\cos 2\phi\right]\,. \tag{3.79}$$

A solution of this equation is given by

$$x_1 = \frac{3G^2M^2}{L^2}\left[\left(1+\frac{1}{2}e^2\right) + e\phi\sin\phi - \frac{1}{6}e^2\cos 2\phi\right] \tag{3.80}$$

as can be verified explicitly by inserting it into (3.79). The first term is a constant offset, while the last term oscillates around zero. The only contribution to the GR perturbation x_1 that increases over time is the middle term. As we only care about this effect, we only keep that term and add it to the Newtonian solution,

$$x = 1 + e\cos\phi + e\frac{3G^2M^2}{L^2}\phi\sin\phi \simeq 1 + e\cos\left[(1-\alpha)\phi\right]\,. \tag{3.81}$$

The second expression is equivalent to the first to first order in the small parameter α, defined as

$$\alpha \equiv \frac{3G^2M^2}{L^2}\,. \tag{3.82}$$

Equation (3.81) corresponds to an ellipse with a period that is not quite 2π, which leads to a slow rotation of the ellipse. The rotation angle per orbit (the perihelion advance) is

$$\Delta\phi = 2\pi\alpha = \frac{6\pi G^2M^2}{L^2}\,. \tag{3.83}$$

In general it is preferable to convert the angular momentum L to another variable that is close to astronomical observations, e.g. the semi-major axis a. Since the perihelion advance of planets in the solar system is small, we can do this with the help of the Newtonian solution (3.77). Geometrically this is an ellipse, for which a is given by

$$a_x = \frac{1}{1-e^2}\,. \tag{3.84}$$

From our definition of x in terms of r, Eq. (3.72) we see that the size of the ellipse was rescaled by a factor of GM/L^2, so that in the original units

$$L^2 = GM\left(1-e^2\right)a \tag{3.85}$$

for the Newtonian case. This is a good enough approximation for us. Including also the speed of light explicitly we find for the perihelion advance finally

$$\Delta\phi = \frac{6\pi GM}{c^2(1-e^2)a}\,. \tag{3.86}$$

Specifically for Mercury this is

$$\Delta\phi(☿) = (6\pi)(1.48\times 10^3\text{m})\frac{1}{(1-0.2056^2)5.79\times 10^{10}\text{m}} = 5.01\times 10^{-7}\text{rad/orbit}\,. \tag{3.87}$$

Mercury orbits the sun once every 88 days, so that in a century the perihelion shift is (in arc seconds)

$$\Delta\phi(☿) = 43.0''/\text{century}\,. \tag{3.88}$$

This corresponds exactly to the anomalous precession that remained after other known factors were corrected for. The natural explanation of the perihelion precession of Mercury was a major success for GR.

The perihelion shift of Mercury is actually an 'essentially' non-linear feature of GR. If we had computed it in the linear weak field limit used the previous chapter, we would have obtained the wrong answer.

3.4.2 Gravitational redshift

An important outcome of the 'Gedanken' experiments of Einstein that motivated the development of GR was the existence of a gravitational redshift due to the equivalence principle, as discussed at the beginning of these lectures. The Schwarzschild metric should also be a good approximation of the exterior gravitational field of the Earth, so that it makes sense to study the gravitational redshift using the Schwarzschild metric.

We consider an observer with four-velocity u^μ, who is at rest with respect to the Schwarzschild coordinate system so that $u^i = 0$. The normalisation $u_\mu u^\mu = g_{\mu\nu}u^\mu u^\nu = -1$ then implies

$$u^0 = \left(1 - \frac{2GM}{r}\right)^{-1/2} . \tag{3.89}$$

If this observer receives a photon propagating along a null geodesic $x^\mu(\lambda)$, she would measure a photon frequency of

$$\omega = -g_{\mu\nu}u^\mu \frac{dx^\nu}{d\lambda} , \tag{3.90}$$

as we already mentioned in (1.192) for Special Relativity, except that now the metric is the Schwarzschild metric, not Minkowski. This implies that

$$\omega = \left(1 - \frac{2GM}{r}\right)^{1/2} \frac{dt}{d\lambda} = \left(1 - \frac{2GM}{r}\right)^{-1/2} E \tag{3.91}$$

where we used the conserved quantity E defined in (3.59). As E is constant along photon trajectories, the photon frequency will necessarily change when measured by observers at rest at different r. For a photon emitted at r_1 and measured at r_2,

$$\frac{\omega_2}{\omega_1} = \left(\frac{1 - \frac{2GM}{r_1}}{1 - \frac{2GM}{r_2}}\right)^{1/2} . \tag{3.92}$$

This is the exact result for the frequency shift in a spherically symmetric space-time. In the limit $r \gg 2GM$, certainly a good approximation on Earth, the corresponding redshift is

$$z \equiv \frac{\lambda_2 - \lambda_1}{\lambda_1} = \frac{\omega_1}{\omega_2} - 1 \simeq \frac{GM}{r_1} - \frac{GM}{r_2} = \phi_2 - \phi_1 = \Delta\phi , \tag{3.93}$$

i.e. the redshift is given by the difference in the gravitational potential. We see that if the photon 'climbs' out of a gravitational potential well ($r_2 > r_1$), the redshift z is positive, corresponding to a reduced frequency (and thus energy) of the photon. For the setup discussed in Section 1.1.3, a static gravitational field with constant gravitational acceleration g, we have that $z = \Delta\phi = gh$, in agreement with Eq. (1.11), once we restore the appropriate factors of c.

3.5 The Schwarzschild black hole

3.5.1 Interior solutions for stars

We now turn to the region $r \leq 2GM$. But before we look for new coordinates that are regular across $r = 2GM$, we want to look at 'interior solutions' of the Schwarzschild metric. With this we mean static and spherically symmetric solutions with a perfect fluid. The perfect fluid is supposed to mimic the interior of stars. We will see that there are limits on how massive stars can be once they are no longer supported by their radiation pressure. Objects that are too heavy will collapse beyond $r = 2GM$, so that in GR it is important to study the end-stadium of such a collapse, i.e. the structure of the space-time left behind in such an event.

We return to our static, spherically symmetric metric (3.32),

$$ds^2 = -f(r)dt^2 + h(r)dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2) = -e^{2\alpha(r)}dt^2 + e^{2\beta(r)}dr^2 + r^2 d\Omega^2\,, \tag{3.94}$$

where $\alpha(r)$ and $\beta(r)$ are now new, unknown functions that need to be determined from the solutions of the full Einstein equations,

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = 8\pi G T_{\mu\nu}\,. \tag{3.95}$$

Using the Ricci tensor (3.35-3.38) and the resulting Ricci scalar

$$R = -2e^{-2\beta}\left[\partial_r^2\alpha + (\partial_r\alpha)^2 - \partial_r\alpha\partial_r\beta + \frac{2}{r}(\partial_r\alpha - \partial_r\beta) + \frac{1}{r^2}\left(1 - e^{2\beta}\right)\right] \tag{3.96}$$

we obtain the Einstein tensor

$$G_{tt} = \frac{1}{r^2}e^{2(\alpha-\beta)}\left(2r\partial_r\beta - 1 + e^{2\beta}\right)\,, \tag{3.97}$$

$$G_{rr} = \frac{1}{r^2}\left(2r\partial_r\alpha + 1 - e^{2\beta}\right)\,, \tag{3.98}$$

$$G_{\theta\theta} = r^2 e^{-2\beta}\left[\partial_r^2\alpha + (\partial_r\alpha)^2 - \partial_r\alpha\partial_r\beta + \frac{1}{r}(\partial_r\alpha - \partial_r\beta)\right] \tag{3.99}$$

$$G_{\phi\phi} = \sin^2\theta G_{\theta\theta}\,. \tag{3.100}$$

As mentioned, we model the star as a perfect fluid, with stress-energy tensor (1.127),

$$T_{\mu\nu} = \rho u_\mu u_\nu + P(g_{\mu\nu} + u_\mu u_\nu) \tag{3.101}$$

To be compatible with the static symmetry of the space-time we take the fluid 4-velocity u^μ pointing in the same direction as the static Killing vector field K^μ, to satisfy the normalisation $u_\mu u^\mu = -1$ it is then given by

$$u_\mu = (e^\alpha, 0, 0, 0)\,. \tag{3.102}$$

Thus in our basis the stress-energy tensor is diagonal, with entries

$$T_{\mu\nu} = \mathrm{diag}\left(e^{2\alpha}\rho, e^{2\beta}P, r^2P, r^2\sin^2\theta P\right)\,. \tag{3.103}$$

We therefore have three independent Einstein equations,

$$\frac{1}{r^2}e^{-2\beta}\left(2r\partial_r\beta - 1 + e^{2\beta}\right) = 8\pi G\rho\,, \tag{3.104}$$

$$\frac{1}{r^2}e^{-2\beta}\left(2r\partial_r\alpha + 1 - e^{2\beta}\right) = 8\pi GP\,, \tag{3.105}$$

$$e^{-2\beta}\left[\partial_r^2\alpha + (\partial_r\alpha)^2 - \partial_r\alpha\partial_r\beta + \frac{1}{r}(\partial_r\alpha - \partial_r\beta)\right] = 8\pi GP\,, \tag{3.106}$$

since the $\phi\phi$ equation is proportional to the $\theta\theta$ equation. The first equation only depends on $\beta(r)$. Replacing $\beta(r)$ through

$$e^{2\beta(r)} = \left(1 - \frac{2Gm(r)}{r}\right)^{-1} \tag{3.107}$$

we find that the tt Einstein equation now is

$$\frac{dm}{dr} = 4\pi r^2 \rho \,, \tag{3.108}$$

so that $m(r)$ is given by

$$m(r) = 4\pi \int_0^r \rho(r') r'^2 dr' \,. \tag{3.109}$$

For a static metric, Σ has to be space-like, implying $e^{2\beta} \geq 0$ or $r \geq 2Gm(r)$.

The rr component of the metric

$$ds^2 = -e^{2\alpha(r)} dt^2 + \left(1 - \frac{2Gm(r)}{r}\right)^{-1} dr^2 + r^2 d\Omega^2 \,, \tag{3.110}$$

can be seen as a generalisation of the Schwarzschild metric, where the mass $m(r)$ can depend on radius. If the radius of star is $R < \infty$, with $\rho = 0$ for $r > R$, then the metric should match the Schwarzschild case at $r = R$,

$$M = m(R) = 4\pi \int_0^R \rho(r) r^2 dr \,. \tag{3.111}$$

M looks like the total mass in a Newtonian setting. However, the proper volume element on Σ is actually $\sqrt{g^{(3)}} d^3x = e^{\beta} r^2 \sin\theta dr d\theta d\phi$, so that the total proper mass is

$$\tilde{M} = 4\pi \int_0^R \rho(r) r^2 e^{\beta(r)} dr = 4\pi \int_0^R \rho(r) \left(1 - \frac{2Gm(r)}{r}\right)^{-1/2} r^2 dr \,. \tag{3.112}$$

The difference between the two is the gravitational binding energy

$$E_B = \tilde{M} - M > 0 \,, \tag{3.113}$$

which is the energy needed to disperse the matter that makes up the star to spatial infinity where the metric is asymptotically flat.

The rr Einstein equation is, using $m(r)$,

$$\frac{d\alpha}{dr} = \frac{Gm(r) + 4\pi G r^3 P}{r[r - 2Gm(r)]} \,. \tag{3.114}$$

In the Newtonian limit where $r^3 P \ll m$ and $Gm(r) \ll r$ we have that

$$\frac{d\alpha}{dr} \simeq \frac{Gm(r)}{r^2} \,. \tag{3.115}$$

This is actually just the once-integrated spherically symmetric Poisson equation $r^{-2}\partial_r(r^2 \partial_r \phi) = 4\pi G \rho$ for the Newtonian gravitational potential ϕ. We can thus consider α as a relativistic analogue of the Newtonian potential.

While we could use the $\theta\theta$ Einstein equation at this point, it is easier to use the conservation equation for the stress-energy tensor, $\nabla_\mu T^{\mu\nu} = 0$. We have already computed the equation for a perfect fluid in Eq. (1.130), for the r component we find

$$(\rho + P)\frac{d\alpha}{dr} = -\frac{dP}{dr} \,. \tag{3.116}$$

Inserting this equation into Eq. (3.114) we find the *Tolman-Oppenheimer-Volkoff (TOV) equation* or *equation of hydrostatic equilibrium*,

$$\frac{dP}{dr} = -(\rho + P)\frac{Gm(r) + 4\pi G r^3 P}{r[r - 2Gm(r)]}\,. \tag{3.117}$$

With (3.109) providing the link between $m(r)$ and ρ we need an additional equation, an equation of state, linking ρ and P to close the system of equations. Typically this will be an equation linking three thermodynamic potentials, e.g. $P(\rho, S)$ where S is the entropy, but we can often just use $P = P(\rho)$. A typical form used in astrophysics is a polytropic equation of state, $P = K\rho^\gamma$ for constants K and γ.

A simple model to gain some insight into key properties of stars is to assume that they are incompressible, so that ρ is constant inside the star, $\rho = \rho_0$ for $r < R$ (and $\rho = 0$ for $r > R$). Having specified a function $\rho(r)$ we can then compute P from Eq. (3.117), i.e. an equation of state is implicitly given. For this case, the function $m(r)$ is

$$m(r) = \begin{cases} \frac{4}{3}\pi r^3 \rho_0\,, & r < R \\ \frac{4}{3}\pi R^3 \rho_0 = M\,, & r > R\,. \end{cases} \tag{3.118}$$

The equation for hydrostatic equilibrium can be integrated exactly in this case, giving

$$P(r) = \rho_0 \left[\frac{R\sqrt{R - 2GM} - \sqrt{R^3 - 2GMr^2}}{\sqrt{R^3 - 2GMr^2} - 3R\sqrt{R - 2GM}}\,,\right] \tag{3.119}$$

and from (3.114) we find that

$$e^{\alpha(r)} = \frac{3}{2}\left(1 - \frac{2GM}{r}\right)^{1/2} - \frac{1}{2}\left(1 - \frac{2GMr^2}{R^3}\right)^{1/2}\,, \quad r < R\,. \tag{3.120}$$

As expected we recover the Schwarzschild metric for $r \to R$.

Equation (3.119) has an interesting property: the central pressure $P(r = 0)$ diverges as the mass approaches

$$M_{\text{max}} = \frac{4}{9G}R\,. \tag{3.121}$$

If a star shrinks beyond this limit, then it cannot remain static and will collapse without limit. Although we derived this limit here using the assumption of a constant density, one can show that it is a fairly general limit. We can rephrase it to say that maximum possible mass of a star of uniform density ρ_0 is

$$M_{\text{max}} = \frac{4}{9(3\pi)^{1/2}}(G^3\rho_0)^{-1/2}\,. \tag{3.122}$$

For a (non-rotating) star of (constant) nuclear density of $\sim 2.3 \times 10^{17}\text{kg}/\text{m}^3$ this maximal mass is about $7.5M_\odot$.

Of course this is a too simplistic model of the actual late-time stage of stellar evolution. In reality the maximal masses will be somewhat lower. Figure 3.3 shows the relation between the core density of stellar remnants and their mass. Very light objects that never become 'real stars' (i.e. able to sustain nuclear fusion of hydrogen to helium) like planets and brown dwarves, with masses well below a solar mass, lie to the left of the diagram. They can be supported basically forever by gas pressure alone, their density is very low and they lie far away from the TOV limit.

Stars on the other hand are supported during their 'normal' life by their radiation pressure, and they can be very heavy (up to hundreds of solar masses – then their own radiation pressure will blow the outer shells away). When they run out of nuclear fuel, the remnant will start being crushed by its own gravitation. For

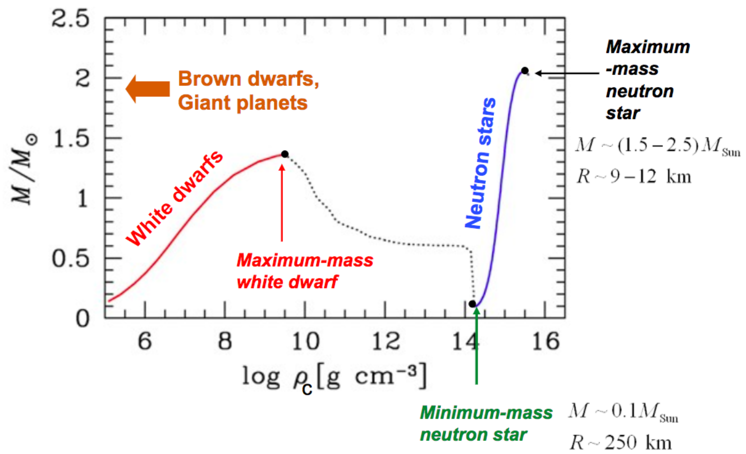

Figure 3.3: A diagram showing the relation between mass and density for stellar remnants. (Origin unknown.)

not so heavy stars (like the sun), the process continues until the electrons start to resist further compression because of the Pauli exclusion principle. Such remnants, supported by electron degeneracy pressure, are called *white dwarves*. They are the most common end-points of stellar evolution, also the sun will most likely end up as a white dwarf in a few billion years. White dwarves are about Earth-sized.

However, as the mass of a white dwarf increases, it shrinks and the electrons become more and more relativistic. At about 1.4 $M_\odot$ it reaches the *Chandrasekhar limit* where the electrons are no longer able to sustain the self-gravitation of the white dwarf. Heavier stars therefore cannot become white dwarves but instead collapse further (although if a white dwarf accretes material until its mass exceeds the Chandrasekhar limit, it may explode like a huge nuclear fusion bomb, a type Ia supernova, rather than collapse).

In remnants with masses beyond the Chandrasekhar limit the electrons combine with protons to form neutrons and neutrinos through electron capture (inverse beta decay). The neutrinos escape, and we are left with a *neutron star*, a kind of huge atomic nucleus[4] with a typical radius of some 10 km. Their density is of the order of nuclear density, $\sim 10^{17}\text{kg}/\text{m}^3$, the density used in our estimate of M_{max} above. We usually see neutron stars with the help of radio telescopes if they are pulsars, or in X-ray if they accrete matter in binary systems. In the future gravitational waves from neutron star mergers, like GW170817, may become a key source of information about them.

Since we do not understand the inner structure and the equation of state of neutron stars well, we do not know their upper mass limit (the *Oppenheimer-Volkoff limit*) from the TOV equation. Very likely they cannot be heavier than about $2 - 2.5M_\odot$, at least we have not observed any neutron stars with masses significantly above $2M_\odot$. Beyond this maximal mass, the stellar remnant has to collapse further. It might end up as an even more compact object composed of a material currently unknown, maybe a quark-gluon plasma that includes strange and even heavier quarks. However, the TOV limit (3.122) decreases as ρ increases, and so very soon such remnants will run out of options. Any objects that are too heavy at the end of their life as stars (and after shedding significant mass during the final stages of their evolution) cannot remain stable at any size and have to shrink beyond $R < 9/4GM$. At this point nothing can stop their collapse, and we end up with an object entirely contained within the Schwarzschild radius $R_S = 2GM$, a *black hole*, giving us finally access to this part of the Schwarzschild metric.

Here we only considered the spherically symmetric situation. In general we expect a collapsing star to retain a certain amount of angular momentum. Originally it was not clear whether this could significantly

[4] Actually the detailed structure is currently not well understood and mostly based on mathematical models. At the surface there is probably a lattice of normal nuclei with an electron liquid, while toward the centre the material will be more and more dominated by free neutrons, maybe forming a superfluid or possibly even something more exotic.

alter the general picture and prevent the formation of black holes. In 1965 Penrose managed to show that the collapse of matter to a singularity is inevitable after a horizon forms, under some mild conditions (for example that the matter has a positive energy density) even without spherical symmetry. He was awarded half of the Nobel Prize of Physics in 2020 for this discovery.

3.5.2 The Kruskal extension

To start with, let us look at radial null curves ($ds^2 = 0$, and θ and ϕ constant) for the Schwarzschild metric (3.47), for which we find

$$\frac{dt}{dr} = \pm\left(1 - \frac{2GM}{r}\right)^{-1} . \tag{3.123}$$

We can think of dt/dr as the slope of light-cones in a $t - r$ diagram. Far away where the Schwarzschild metric becomes asymptotically flat, the light-cone slope is ± 1, as expected for flat space. But as we approach $r \to R_S = 2GM$ we find that $dt/dr \to \infty$, i.e. the light-cones 'close up', implying that a light-ray approaching R_S will never actually get there, at least not for finite coordinate time t.

However, let us return to the equation of motion of a test-body, Eq. (3.63), with $L = 0$ for simplicity. For a massive observer with $\epsilon = 1$, the affine parameter λ is proper time τ so that

$$\dot{r}^2 = \frac{2GM}{r} + E^2 - 1\,. \tag{3.124}$$

Starting at rest from an initial position at $r = R$, for which $2GM/R = 1 - E^2$ and $E < 1$ we have

$$d\tau = \left(\frac{2GM}{r} - \frac{2GM}{R}\right)^{-1/2} dr\,, \tag{3.125}$$

which is solved by a cycloid, with parametric representation

$$r = \frac{R}{2}(1 + \cos\eta) \tag{3.126}$$

$$\tau = \sqrt{\frac{R^3}{8GM}}(\eta + \sin\eta)\,. \tag{3.127}$$

At $\eta = 0$ we are at the starting point $\tau = 0$ and $r = R$, and we reach $r = 0$ at $\tau = (\pi/2)\sqrt{R^3/2GM}$, which is the proper time for free fall to the origin. Nothing special happens at $r = 2GM$. But if we computed r in terms of coordinate time t, using

$$\dot{r} = \frac{dr}{dt}\dot{t} = \frac{dr}{dt}\frac{E}{1 - 2GM/r}\,, \tag{3.128}$$

we would find that r approaches $r = 2GM$ only asymptotically for $t \to \infty$, as expected from the light-cone argument above.

In other words, an observer falling into a black hole reaches $r = 0$ in finite proper time. But a second observer far away, with a proper time that is given roughly by coordinate time, would never see the infalling observer cross the Schwarzschild radius. Instead the first observer would slow down as she approaches that radius. The distant observer would then actually quickly lose sight of the infalling observer, due to the gravitational redshift of the photons, but still the infalling observer would apparently never cross $r = R_S$.

As the infalling observer has no problems reaching $r < R_S$ in finite proper time, it seems that our coordinate system may not be optimally chosen to describe the black hole. We follow [4] in trying to motivate a better choice. We start by fixing the problem with the narrowing light-cones. Integrating (3.123) we obtain

$$t = \pm r^* + \text{const.} \tag{3.129}$$

for the *tortoise coordinate* r^*

$$r^* \equiv r + 2GM \ln\left(\frac{r}{2GM} - 1\right), \quad r > 2GM\,. \tag{3.130}$$

By construction the light-cones have slope ± 1 in coordinates (t, r^*), and the Schwarzschild metric is given by

$$ds^2 = \left(1 - \frac{2GM}{r}\right)(-dt^2 + dr^{*2}) + r^2 d\Omega^2\,. \tag{3.131}$$

However, what we really did was to push $r = 2GM$ to infinity.

We now introduce light-cone coordinates

$$v \equiv t + r^*\,, \quad u \equiv t - r^*\,, \tag{3.132}$$

and we express the metric in terms of the original radial coordinate r but replace t with v. This choice is known as *Eddington-Finkelstein coordinates*. The metric then becomes

$$ds^2 = -\left(1 - \frac{2GM}{r}\right)dv^2 + (dvdr + drdv) + r^2 d\Omega^2\,. \tag{3.133}$$

Now the metric is regular at $r = 2GM$, with determinant $g = -r^4 \sin^2\theta$, i.e. it can be inverted. We can conclude now that $r = 2GM$ is really just a coordinate singularity in the Schwarzschild coordinates. In Eddington-Finkelstein coordinates the light-cones are determined by

$$\frac{dv}{dr} = \begin{cases} 0 & \text{(inward pointing)} \\ 2\left(1 - \frac{2GM}{r}\right)^{-1} & \text{(outward pointing)} \end{cases} \tag{3.134}$$

In this coordinate system the part of the light-cone pointing toward the black hole is always aligned with the v constant axis, while the direction pointing ‘away’ from the black hole is tilting over. At $r = 2GM$ the light-cones are well-behaved, but the outward pointing direction is now pointing along constant r, and for $r < 2GM$ the whole future light-cone is pointing toward smaller r. The surface $r = 2GM$ functions therefore as a ‘surface of no return’, an *event horizon* (in the sense that it is impossible to escape to infinity if one finds oneself on the inside).

As the light-cone is aligned with the $r = 2GM$ event horizon surface, this is actually a null surface even though it is at constant radius. Inside the horizon, the singularity at $r = 0$ is now in the future of an observer, i.e. it becomes an (unavoidable) future event rather than a spatial location. This means that the causal structure of the Schwarzschild space-time itself prevents anything from crossing $r = 2GM$ from the inside to the outside. This impossibility to ‘see inside’ is also the reason for the name *black hole*. We also note that the notion of horizon is a global notion, it is not possible to figure out where a horizon is from just the local geometry. A local observer would not need to realise that she has just crossed an horizon.

But we could have made a different choice above. We could have replaced t with u instead of v. Then the metric would have been

$$ds^2 = -\left(1 - \frac{2GM}{r}\right)du^2 - (dudr + drdu) + r^2 d\Omega^2\,. \tag{3.135}$$

Now the light-cones are tilting in just the opposite direction, i.e. future-directed paths can only cross the horizon from the inside to the outside. This is not actually a contradiction, these two choices describe different extensions of the original metric: The tortoise coordinate r^* goes to $-\infty$ for $r \to 2GM$. If we do this along constant v, as in the first case, then we need to have $t \to \infty$, while for constant u we have that $t \to -\infty$. Therefore the first choice, (v, t), extends the Schwarzschild metric toward the future, while the second choice extends it toward the past.

A good choice of coordinates then might be to use (u, v) instead of (t, r)? The problem is that in these coordinates, $r = 2GM$ is again only reachable for either $u \to \infty$ or $v \to \infty$, since

$$\frac{1}{2}(v-u) = r^* = r + 2GM \ln\left(\frac{r}{2GM} - 1\right) . \tag{3.136}$$

But we can use the transformed coordinates

$$v' \equiv e^{v/4GM} , \quad u' \equiv -e^{-u/4GM} \tag{3.137}$$

to solve this problem. In terms of the original coordinates (t, r), they are given by

$$v' = \left(\frac{r}{2GM} - 1\right)^{1/2} e^{(r+t)/4GM} , \quad u' = -\left(\frac{r}{2GM} - 1\right)^{1/2} e^{(r-t)/4GM} , \tag{3.138}$$

and the Schwarzschild metric in the (v', u', θ, ϕ) coordinate system is

$$ds^2 = -\frac{16G^3M^3}{r} e^{-r/2GM}(dv'du' + du'dv') + r^2 d\Omega^2 . \tag{3.139}$$

At the horizon $r = 2GM$, none of the metric coefficients exhibit any strange behaviour.

Although there is no problem using a coordinate system with two null and two space-like basis vectors, it is conventional to transform u' and v' back to a time-like and a space-like coordinate, by defining

$$T \equiv \frac{1}{2}(v' + u') = \left(\frac{r}{2GM} - 1\right)^{1/2} e^{r/4GM} \sinh\left(\frac{t}{4GM}\right) , \tag{3.140}$$

$$R \equiv \frac{1}{2}(v' - u') = \left(\frac{r}{2GM} - 1\right)^{1/2} e^{r/4GM} \cosh\left(\frac{t}{4GM}\right) \tag{3.141}$$

for the region where the Schwarzschild coordinates were valid, i.e. for $r > 2GM$. In the (T, R) coordinates, the metric becomes

$$ds^2 = \frac{32G^3M^3}{r} e^{-r/2GM}(-dT^2 + dR^2) + r^2 d\Omega^2 , \tag{3.142}$$

where r is given implicitly by

$$T^2 - R^2 = \left(1 - \frac{r}{2GM}\right) e^{r/2GM} . \tag{3.143}$$

The coordinates (T, R, θ, ϕ) are called *Kruskal coordinates*, or Kruskal-Szekeres coordinates. They represent a very convenient choice of coordinates for several reasons. Firstly, in the two-dimensional submanifold $\theta = \phi = \text{const.}$ the metric is conformally equivalent to the Minkowksi metric, and the radial null-curves thus look like in flat space,

$$T = \pm R + \text{const.} , \tag{3.144}$$

i.e. the light-cones do not contract. The event horizon at $r = 2GM$ is however still at a finite distance, in the new coordinates it is given by

$$T = \pm R . \tag{3.145}$$

We see now explicitly that the event horizon is a null surface. More generally, the surfaces of $r = \mathrm{const.}$ correspond to, from (3.143),

$$T^2 - R^2 = \mathrm{const.}\,, \tag{3.146}$$

i.e. they are parabolas in the R-T plane, with the event horizon as the limiting case of straight lines through the origin. The surfaces of constant t on the other hand lie at

$$\frac{T}{R} = \tanh\left(\frac{t}{4GM}\right) \tag{3.147}$$

from the ratio of (3.140) and (3.141). These are straight lines through the origin with slope $\tanh(t/4GM)$. We see explicitly that the surface $r = 2GM$ corresponds to $t = \pm\infty$ and that the region $r < 2GM$ is not accessible in our old coordinates for real t.

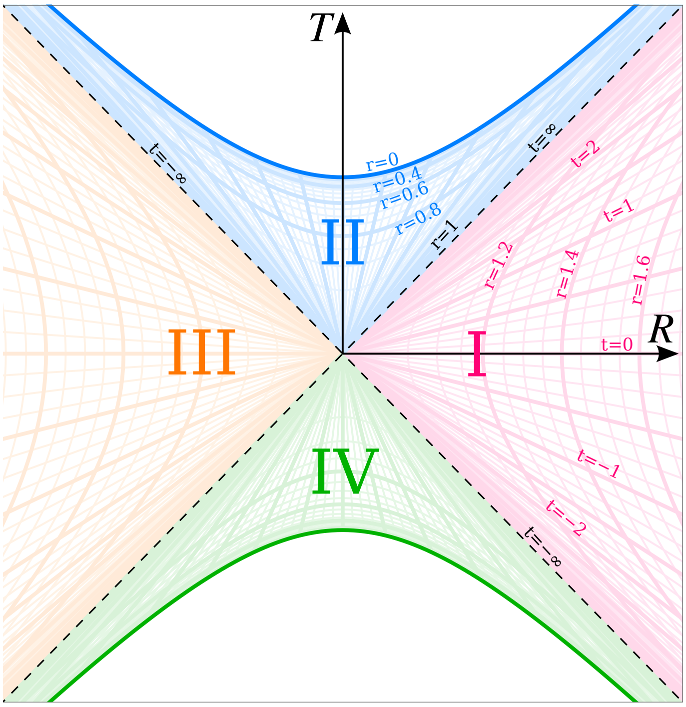

Figure 3.4: A Kruskal diagram. Image from Wikipedia (Dr. Greg, CC BY-SA 3.0 license, replaced X by R, removed some labels).

But the metric (3.142) with r defined implicitly through (3.143) is regular in a larger region than the original Schwarzschild coordinates. In Kruskal coordinates, (T, R) can take any value outside of the true singularity at $r = 0$, i.e.

$$-\infty \leq R \leq \infty\,, \quad T^2 < R^2 + 1\,. \tag{3.148}$$

We can draw a space-time diagram in the (T, R) plane to summarise the discussion above, shown in Fig. 3.4. Each point in the figure corresponds to a two-sphere as the coordinates (θ, ϕ) are suppressed. This diagram shows the maximal extension of the Schwarzschild geometry, in the sense that every geodesic can either be extended to arbitrarily large values of the affine parameter or else hits the singularity at $T^2 - R^2 = 1$[5] for a finite value of the affine parameter. The Ricci tensor vanishes everywhere on this *Schwarzschild-Kruskal manifold*. The manifold is conventionally split into four regions. Region I corresponds to the region where the original coordinates (t, r) are well behaved, $r > 2GM$. From region I we can reach

[5] This is a true singularity since, as we have seen, e.g. the Kretschmann scalar $R_{\mu\nu\lambda\rho}R^{\mu\nu\lambda\rho}$ diverges there.

region II along future-directed null rays, while region IV is connected to region I via past-directed null rays, i.e. region I is in the future of region IV.

Region II is what we call the black hole. From inside region II there is no future-directed way back out. Actually all future-directed paths in region II will reach the singularity at $r = 0$, as we already mentioned. Region IV is the time-reverse of region II that we found above when we used (u, r) instead of (v, r), the two regions are isometric to each other. Things can move from region IV to region I, but never from region I to region IV. This is often called a 'white hole' as the opposite of a black hole. Time-like trajectories in region IV have a singularity in their past (just as time-like trajectories in region II have a singularity in their future). Region III is an isometric copy of region I, but it is unreachable from region I, and region I cannot be reached from region III. Region I and III are connected by an *Einstein-Rosen bridge* or wormhole but the wormhole cannot be traversed. While the Schwarzschild-Kruskal manifold is static (by construction) in regions I and III, it is dynamical in regions II and IV, in the sense that the Killing vector field $\mathbf{K} = \boldsymbol{\partial}_t$ becomes space-like, so that there are no observers at rest in regions II and IV. (At the horizon, $\mathbf{K}$ is null, $g_{\mu\nu}K^\mu K^\nu = 2GM/r - 1 = 0$.)

Formation of a black hole by stellar collapse

Let's follow schematically the evolution of a star that collapses after crossing the Oppenheimer-Volkoff limit, e.g. after running out of nuclear fuel or accreting mass from a companion star. A sketch is shown in Figure 3.5. As the stellar radius becomes smaller than its own Schwarzschild radius, equilibrium becomes obviously impossible as all future-directed geodesics are now pointing 'inward'. Collapse to a singularity is then unavoidable. Also light signals from the star can no longer reach any distant observer, the final collapse of the star and the formation of the singularity are 'hidden' from outside view [6]. However, an observer sitting on the star may not immediately notice anything special as the horizon is crossed, especially for a very massive star where the tidal forces at horizon crossing can be small. An external observer however would see the shrinking of the star slowing down as it approaches the forming horizon. A detailed calculation shows that the redshift is increasing exponentially fast close to the horizon, with typical time-scale of $\tau \sim R_S/c \simeq 10 M/M_\odot \mu$s, so that in reality a collapsing star would seem to suddenly vanish – an astrophysical black hole does appear black.

While the Kruskal diagram, Fig. 3.4, shows the maximal extension of the Schwarzschild manifold, the full extension is not relevant in astrophysical situations. As we saw, inside a (spherically symmetric) star, the metric is not described by the Schwarzschild metric. It is not a vacuum solution, the metric is regular, there are no singularities or horizons hidden there (until the star creates its own horizon during the collapse). This effectively removes regions III and IV from the Kruskal diagram. There is no white hole in the past of a collapsing star, and no Einstein-Rosen bridge.

Spaghettification

For 'fun' we can also compute tidal forces that are experienced by matter (or observers) falling into the black hole. An in-falling observer with 4-velocity $\mathbf{u}$ would use his own rest-frame as his reference frame. We can compute the Riemann tensor in this frame by performing a Lorentz-boost to the observer frame. It turns out that the components of the Riemann tensor are invariant under such a boost. Another aspect of the Schwarzschild metric that simplifies our life is that inside the horizon t and r change roles, but the Riemann tensor remains the same.

We write the geodesic deviation equation (1.106) for two freely falling objects separated by a vector $\mathbf{X}$ as

$$\ddot{X}^i = R^i_{00j} X^j \,, \tag{3.149}$$

[6]The *cosmic censorship* conjecture says that all singularities are hidden behind a horizon and cannot be directly observed.

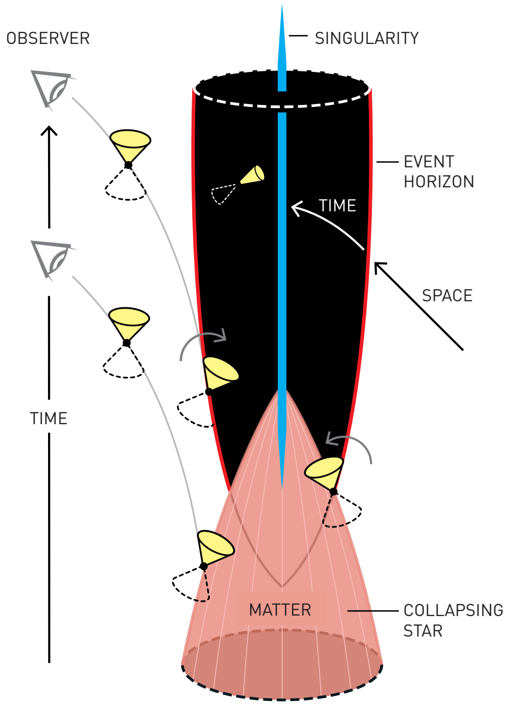

Figure 3.5: A sketch of the formation of a black hole from stellar collapse, from the 2020 Nobel Prize in Physics press release. Copyright Johan Jarnestad/The Royal Swedish Academy of Sciences.

where the dot is derivation with respect to proper time. Using the Riemann tensor (3.34) for the Schwarzschild metric (3.47) in a Cartesian coordinate system that is falling along the $i = 1$ direction toward a black hole, we find

$$\ddot{X}^1 = \frac{2GM}{r^3} X^1 \,, \quad \ddot{X}^2 = -\frac{GM}{r^3} X^2 \,, \quad \ddot{X}^3 = -\frac{GM}{r^3} X^3 \,. \tag{3.150}$$

We consider an observer with height l (along the 1 direction), width w (along the 2 and 3 directions) and weight μ is falling head-first into the black hole. For the time being we assume that the observer remains intact and we want to know the forces acting on him. From (3.150) we see that the relative acceleration of a body-part of the observer a distance h from the center of mass is $a = (2GM/r^3)h$, and the resulting force is therefore

$$dF = a d\mu = \frac{2GM}{r^3} h d\mu \,. \tag{3.151}$$

The total force on the cross-sectional plane through the center of mass is then

$$F = \int_0^{l/2} \frac{2GMh}{r^3} \frac{\mu}{lw^2} w^2 dh = \frac{1}{4} \frac{\mu GMl}{r^3} \,. \tag{3.152}$$

The longitudinal stress or force per area is $T^l = -F/w^2$ so that for an observer with $\mu = 75\text{kg}$, $l = 1.8\text{m}$ and $w = 0.2\text{m}$

$$T^l = -\frac{1}{4} \frac{\mu GMl}{r^3 w^2} \simeq -1.1 \times 10^{14} \frac{M}{M_\odot} \left(\frac{r}{1\text{km}} \right)^{-3} \frac{\text{N}}{\text{m}^2} \,, \tag{3.153}$$

and similarly the lateral stresses are

$$T_\perp = \frac{1}{8} \frac{\mu GM}{lr^3} \simeq 0.7 \times 10^{12} \frac{M}{M_\odot} \left(\frac{r}{1\text{km}} \right)^{-3} \frac{\text{N}}{\text{m}^2} \,. \tag{3.154}$$

For comparison, 1 atmosphere is about $10^5\mathrm{N/m}^2$, and the tensile strength of steel is around $10^9\mathrm{N/m}^2$. The good news is that for very massive black holes, the tidal forces at the horizon can be harmless. The bad news is that, eventually, the observer *will* be pulled apart along the longitudinal axis and compressed laterally, a process called *spaghettification*. As [2] puts it, it is thus not advisable to fall into a black hole.

Penrose diagrams

An often-used tool to visualize the structure of a space-time is through a conformal compactification that allows to include also regions at infinity in local calculations and in diagrams like Fig. 3.4. A simple standard example is the compactification of the Minkowski space-time. The metric in polar coordinates is

$$ds^2 = -dt^2 + dr^2 + r^2 d\Omega^2 \,. \tag{3.155}$$

As in (3.132) we introduce light-cone or null coordinates $u = t - r$ and $v = t + r$, in terms of which the metric becomes

$$ds^2 = -du dv + \frac{1}{4}(v-u)^2 d\Omega^2 \,. \tag{3.156}$$

Both coordinates u and v can take any real value, subject to the constraint $v \geq u$ as $r \geq 0$. We can now use a diffeomorphism to map this unbounded range to a bounded region, for example by introducing new coordinates U and V through

$$u = \tan U \,, \quad v = \tan V \,, \tag{3.157}$$

for $U, V \in (-\pi/2, \pi/2)$ and $V - U \geq 0$. Using

$$du dv = \frac{1}{\cos^2 U \cos^2 V} dU dV \,, \quad (v-u)^2 = \frac{1}{\cos^2 U \cos^2 V} \sin^2(U - V) \,, \tag{3.158}$$

we can write for the line element in terms of the coordinates (U, V, θ, ϕ)

$$ds^2 = \frac{1}{4\cos^2 U \cos^2 V} d\tilde{s}^2 \,, \quad d\tilde{s}^2 = -4 dU dV + \sin^2(U - V) d\Omega^2 \,. \tag{3.159}$$

Therefore the line elements and thus the corresponding metric $g_{\mu\nu}$ and $\tilde{g}_{\mu\nu}$ are conformally related,

$$\tilde{g} = \Omega^2 g \,, \quad \Omega^2 = 4\cos^2 U \cos^2 V \,. \tag{3.160}$$

(Note that this Ω^2 is the conventional symbol for the conformal factor and has nothing to do with the angular part of the metric $d\Omega^2 = d\theta^2 + \sin^2\theta d\phi^2$.) The metric $\tilde{g}$ is also well-defined when U and/or V have the value $\pm\pi/2$, i.e. $\tilde{g}$ is a smooth metric on the compact manifold $[-\pi/2, \pi/2]^2 \times S^2$.

We can now re-introduce a time and radial coordinate through $\tilde{t} = U + V$ and $\chi = V - U$, in terms of which

$$d\tilde{s}^2 = -d\tilde{t}^2 + d\chi^2 + \sin^2\chi d\Omega^2 \,. \tag{3.161}$$

The image of the Minkowki space-time is the open part of the compact manifold given above. Its boundary is called *conformal infinity of Minkowski space-time*, it consists of the following pieces:

- The two null-hypersurfaces $\mathscr{I}^+$ and $\mathscr{I}^-$ (pronounced 'scri-plus' and 'scri-minus'), given by $\mathscr{I}^+ = \{V = \pi/2, |U| < \pi/2\}$ and $\mathscr{I}^- = \{U = -\pi/2, |V| < \pi/2\}$. These hypersurfaces represent future and past null infinity, i.e. the places where null-geodesics end.

- The points $i^\pm$, given by $U = V = \pm\pi/2$, they represent future and past time-like infinity, this is where time-like geodesics end.

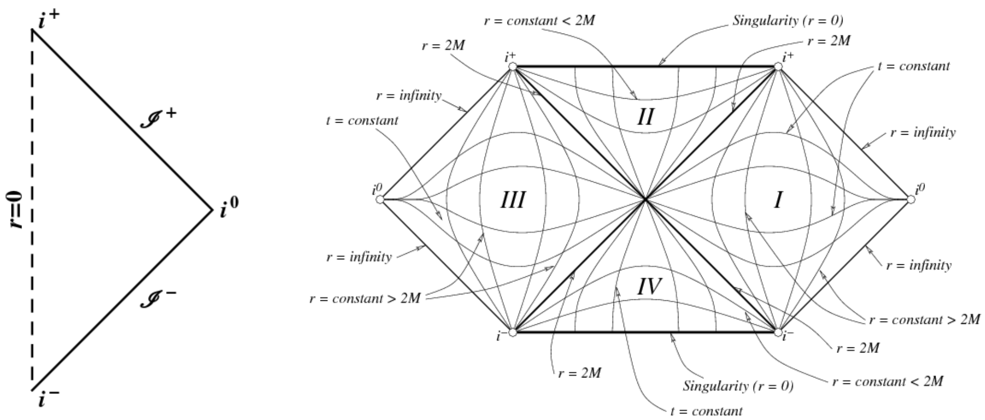

Figure 3.6: Left: Conformal diagram of the Minkowski space-time (from http://physics.oregonstate.edu/coursewikis/GGR/book/ggr/penrose). Right: Penrose diagram for the Schwarzschild-Kruskal space-time (from physics.stackexchange.com).

- The point i^0, corresponding to $U = -\pi/2$, $V = \pi/2$, representing space-like infinity, which is the end of all space-like geodesics.

We show the corresponding diagram in $(\tilde{t}, \chi)$ coordinates in the left panel of Fig. 3.6.

We can proceed analogously for the Schwarzschild-Kruskal space-time. We use the coordinates u' and v' introduced in Eq. (3.137), related to T and R through $u' = T + R$ and $v' = T - R$, and the proceed as before with the transformation 3.157

$$\frac{u'}{\sqrt{2GM}} = \tan U\,, \qquad \frac{v'}{\sqrt{2GM}} = \tan V\,, \tag{3.162}$$

with $(U, V) \in (-\pi/2, \pi/2)$ and $U - V \geq 0$. We also introduce $\tilde{T} = U + V$ and $\tilde{R} = U - V$. We show the corresponding diagram in the right panel of Fig. 3.6.

3.5.3 Observational evidence for black holes

How sure are we that black holes are real? The black holes themselves do not emit radiation (except possibly for Hawking radiation) but matter that falls into them would generally heat up and radiate as it is disrupted (spaghettified), and of course black holes curve the space-time around them, so that they can be seen through their influence on the motion of objects nearby. But the probably most concrete current and especially future evidence comes from the analysis of gravitational wave signals from black hole - black hole mergers, as those can probe the GW emission mechanism, which is a direct observation of a part of the metric during the merger event, as discussed in the weak field limit in Chapter 2. Of course for gravitational waves from the merger of black holes or other very compact objects it is necessary to go beyond the weak field limit, with advanced perturbative techniques and numerical simulations.

The measurements obtained so far agree very well with expectations for black holes in GR, and in the first (and so far strongest) detection, GW150914, the GW signal indicates that the two merging objects were only separated by about 350km at the end, which is about four times the Schwarzschild radius of the inferred masses. In Fig. 2.4 we showed the data and simulations (template) for that event, looking at the waveform we see the increasing oscillations as the two masses approach, and then the peak and decay from the merger and subsequent ring-down of the merged black hole. As more data accumulates, our constraints on the existence and nature of black holes will dramatically improve. Additionally, we would not expect an optical counterpart from the merger of two isolated black holes, while the collision of two super-dense bodies moving at a good fraction of the speed of light should give rise to some pretty spectacular fireworks – as probably observed for the combined event GW170817 (gravitational wave signal detected with LIGO and Virgo), GRB 170817A (gamma ray burst detected by the Fermi and INTEGRAL satellites 1.7 seconds after the arrival of the GW signal) and AT 2017gfo (an optical transient found 11 hours later in the galaxy NGC 4993 while observing the area indicated by the GW signal).

Astrophysical black holes often accrete matter from their surroundings, e.g. from a companion star. This matter is expected to accumulate in a disc around the central dark, compact object, due to angular momentum conservation. Friction in the disc allows to redistribute angular momentum, so that some matter can fall inward. During this process, up to 40% of the rest mass of the accreted material can be released as radiation. Due to this very high efficiency, black hole 'engines' are thought be behind many of the most energetic processes in the universe. The first good observational candidate for a black hole, Cygnus X-1, is an X-ray binary, a binary star systems where a black hole accretes material from a second star, while supermassive black holes are thought to power active galactic nuclei (AGN) and quasars.

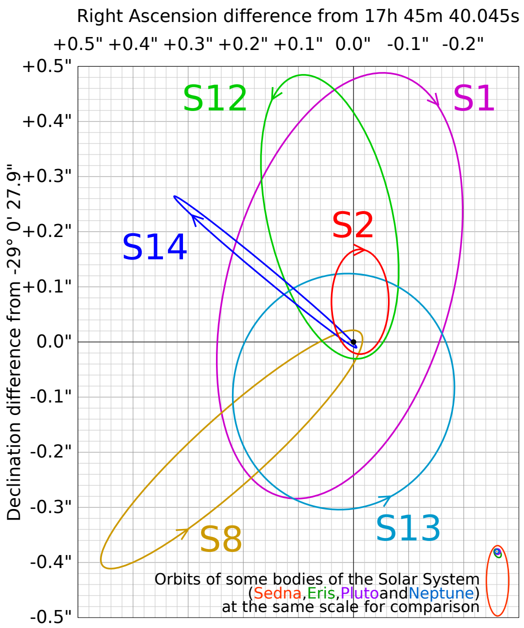

Figure 3.7: Inferred orbits of six stars around supermassive black hole candidate Sagittarius A* at the Milky Way galactic centre based on data from Eisenhauer et al, ApJ **628**, 246 (2005). In 2020 Reinhard Genzel and Andrea Ghez shared half of the Nobel Prize in Physics for these observations. Image from Wikipedia (Cmglee, CC BY-SA 3.0 license, unmodified).

Most, if not all galaxies appear to host supermassive black holes in their center. Also our own galaxy, the Milky Way, contains a very compact and radio-bright object called *Sagittarius A**. Figure 3.7 shows the orbits of stars around this source. The mass inferred from their trajectories is about $4\times10^6 M_\odot$. The star S2 orbits Sagitarrius A* once every 15.2 years, and its closest distance to it is about 1.8×10^{10} km. The closest distance of the star S14 is even only about 6.7×10^9 km. The highest resolution radio measurements of the Milky Way center indicate a diameter of the source of about 37μas, which would correspond to about 44×10^6 km (about twice the horizon size of the black hole). It is difficult to imagine what else rather than a supermassive black hole could explain these observations. Recently the Event Horizon Telescope, a worldwide network of radio telescopes working together to form an 'Earth-sized' interferometric telescope with a resolution of 20μas[7], managed to image what appears to be the gravitationally lensed accretion disk

[7] According to the press-release, a resolution high enough to read a newspaper in New York from Paris.

around a supermassive black hole in the center of the galaxy M87, 55 million light years away, shown in Fig. 3.8.

Mergers of such supermassive black holes cannot be seen with the current ground-based interferometric gravitational wave detectors, as the frequency range is too low. In the future, space-based GW observatories like the ESA LISA satellite project, slated for launch around 2034, should fill this gap.

Figure 3.8: In 2019 the Event Horizon Telescope published this image of a bright ring formed by light bent in the strong gravitational field of what is probably a 6.5×10^9 solar mass black hole in the center of the galaxy M87. The image was obtained by multiple radio telescopes working together to obtain an interferometric image at a wavelength of 1.3mm. It is consistent with expectations for such a black hole. The black hole event horizon would be about 2.5 times smaller than the 'shadow' at the center and would have a diameter of about 40×10^9 km. Image credit: Event Horizon Telescope Collaboration

The observations indicate that there are at least two classes of black holes, 'stellar mass' black holes (with masses up to several tens of solar masses at least), and supermassive black holes with masses in the range of millions to billions of solar masses. According to GR, stellar mass black holes are the effectively inevitable end-product of the evolution of sufficiently massive stars. At this time we have no good understanding of how the supermassive black holes were created, and whether there is also a population of intermediate mass black holes or not.

3.5.4 Beyond classical Schwarzschild black holes

In this final section we briefly discuss some additional results about black holes that go beyond the Schwarzschild solution discussed so far. We will not go into details, more information can be found in most textbooks.

The no-hair theorem

Stars normally rotate, and when they collapse to a black hole we would expect some of the angular momentum to survive. Also, nothing should prevent us from throwing charged particles into a black hole, which then should be charged. However, surprisingly we do not need more information than this to characterise a static black hole. This is the topic of the *no-hair theorem*:

> Stationary, asymptotically flat black hole solutions of GR coupled to electromagnetism that are non-singular outside of the event horizon are fully characterised by the following parameters: mass, electric (and possibly magnetic) charge, and angular momentum.

The stationarity condition may seem restrictive, but simulations indicate that black holes very quickly reach a stationary state after a disturbance (like a merger, or something falling into the black hole), through emission of gravitational waves. The no-hair theorem can be somewhat broken, especially if other long-range fields are added. But generally speaking black holes are very simple objects, characterised by only a few numbers. This can lead to theoretical problems when quantum effects near black hole horizons are considered, as we briefly discuss later on.

Charged and rotating black holes

A charged black hole[8] is described by the *Reissner-Nordström* (RN) metric,

$$ds^2 = -\Delta dt^2 + \Delta^{-1} dr^2 + r^2 d\Omega^2 \,, \quad \Delta = 1 - \frac{2GM}{r} + \frac{G(Q^2 + P^2)}{r^2} \,. \tag{3.163}$$

This is a relatively straightforward extension of the Schwarzschild case as the black hole is still spherically symmetric. M is the mass of the black hole as in Schwarzschild, and Q and P are the total electric and magnetic charge. The event horizon, similarly as for Schwarzschild, is at $\Delta(r) = 0$. But this is now a quadratic equation with solutions

$$r_\pm = GM \pm \sqrt{G^2 M^2 - G(Q^2 + P^2)} \,. \tag{3.164}$$

This can have two, one or zero solutions, depending on the relative magnitude of GM^2 and $Q^2 + P^2$.

If $Q^2 + P^2 > GM^2$ then $\Delta > 0$ and the metric is perfectly regular in (t, r, θ, ϕ) coordinates to $r = 0$, where we have a 'naked' singularity, in the sense that it is not hidden by a horizon. The naked singularity is actually gravitationally repulsive, but null geodesics can reach it. Not only because of the cosmic censorship conjecture, but also because the total energy of the black hole is less than the energy of the electromagnetic field alone, is this solution generally considered to be unphysical.

The case $GM^2 > Q^2 + P^2$ on the other hand describes the situation one would expect in gravitational collapse. In this case there are two null surfaces at $r_\pm$, both are horizons. The outside horizon at r_+ is very similar to the Schwarzschild horizon, with r and t switching roles inside r_+ and with outside observers seeing infalling objects slowing down. But an infalling observer inevitably reaches r_-, where r and t switch back. Now the motion to decreasing r can be stopped, an observer does not need to hit the singularity at $r = 0$, which is a time-like line, not space-like as in Schwarzschild. The observer can even reverse her motion and exit through the r_- horizon. At this point she *has* to move again, now out through the r_+ horizon, which is like emerging from the white hole in Schwarzschild. The observer can then fall again into the black hole (which will be another hole actually), and repeat the process, as illustrated in the left Penrose diagram of Fig. 3.9 (ignore the 'antiverse' bits).

Finally if $GM^2 = Q^2 + P^2$ then we have an *extremal Reissner-Nordström* solution. In this case r_+ and r_- coincide. There is still a horizon at $r_+ = r_-$, but while the r coordinate is a null direction at $r = GM$, it is space-like on either side and an observer does not need to fall into the (time-like) singularity at $r = 0$, and can instead visit again other copies of the 'outside' universe. The conformal diagram is shown in the middle panel of Fig. 3.9.

Rotating black holes are more difficult to describe, as they are not spherically symmetric. For this reason the solution was only found in 1963 by Kerr. The *Kerr metric* is given by

$$\begin{aligned} ds^2 = &-\left(1 - \frac{2GMr}{\rho^2}\right) dt^2 - \frac{2GMar\sin^2\theta}{\rho^2}(dt d\phi + d\phi dt) + \frac{\rho^2}{\Delta} dr^2 \\ &+ \rho^2 d\theta^2 + \frac{\sin^2\theta}{\rho^2}\left[(r^2 + a^2) - a^2 \Delta \sin^2\theta\right] d\phi^2 \,, \end{aligned} \tag{3.165}$$

where

$$\Delta(r) = r^2 - 2GMr + a^2 \,, \quad \rho^2(r, \theta) = r^2 + a^2 \cos^2\theta \,. \tag{3.166}$$

[8] Charged black holes are probably not too important in an astrophysical setting, as a highly charged black hole would create an electric field that attracts opposite charges from matter nearby, so that the excess charge of the black hole would be quickly reduced.

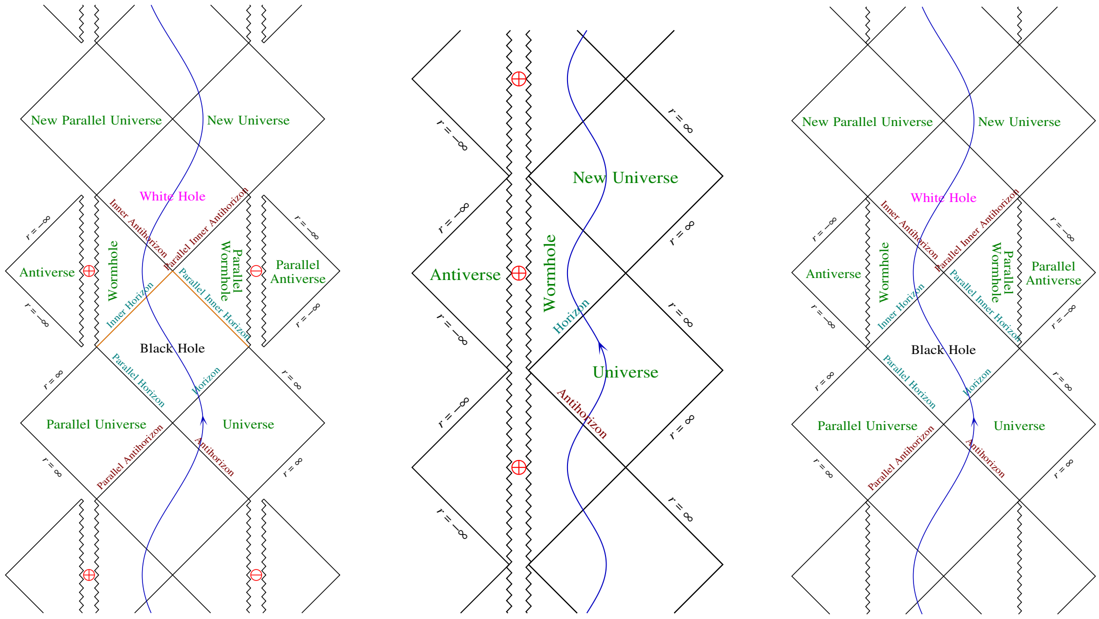

Figure 3.9: Penrose diagrams for Reissner-Nordström black holes with $GM^2 > Q^2 + P^2$ (left), $GM^2 = Q^2 + P^2$ (extremal case, middle) and for a Kerr black hole (right). The blue line shows a possible time-like trajectory of an observer. The oscillatory vertical lines are the singularities. (From https://jila.colorado.edu/ ajsh/insidebh/penrose.html.)

The space of solutions is parameterised by M and a, where M is again the black hole mass and $a = J/M$ is the angular momentum per unit mass. By replacing $2GMr \to 2GMr - G(Q^2 + P^2)$ it is possible to also include charge, this is then the *Kerr-Newmann metric*.

The coordinates (t, r, θ, ϕ) in which we wrote the Kerr metric are known as *Boyer-Lindquist coordinates*. For $a \to 0$ they reduce to the Schwarzschild coordinates. The horizons are again where $g^{rr} = \Delta/\rho^2 = 0$. It is enough that Δ vanishes as $\rho^2 \geq 0$. This happens for

$$r_\pm = GM \pm \sqrt{G^2M^2 - a^2}\,. \tag{3.167}$$

As for RN there are three cases, $GM < a$ (with a naked singularity), the extremal case $GM = a$ and the 'normal' case $GM > a$, which is the case that we will study more closely. Both radii $r_\pm$ are event horizons, a sketch of the situation is shown in Fig. 3.10.

The Kerr space-time is stationary but not static, this can be seen e.g. by noticing that the Kerr metric (3.165) is not invariant under time reversal. Also, there are two Killing vector fields, $\mathbf{K} = \boldsymbol{\partial}_t$ and $\mathbf{R} = \boldsymbol{\partial}_\phi$, but now $\mathbf{K}$ is not orthogonal to the $t = \text{const.}$ hypersurfaces. This is of course no surprise, since the Kerr black hole is rotating.

The Killing vector $\mathbf{K}$ is null, $K^\mu K_\mu = 0$, on the stationary limit surface, given by $g_{tt} = 0$, or

$$(r - GM)^2 = G^2M^2 - a^2\cos^2\theta\,, \tag{3.168}$$

while the outer horizon is according to (3.167) given by

$$(r_+ - GM)^2 = G^2M^2 - a^2\,. \tag{3.169}$$

The region between these two surfaces is called the *ergosphere*. It is possible to leave the ergosphere both through the horizon and through the stationary limit surface (which is not a horizon), but inside the ergosphere it is impossible to remain stationary, any observer has to rotate with the black hole in the ϕ

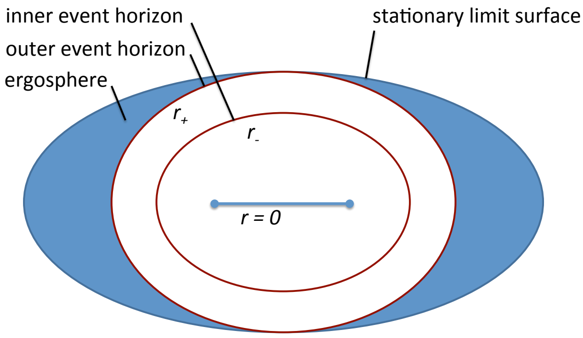

Figure 3.10: The horizon structure of a Kerr black hole. The horizons at r_+ and r_- are null surfaces and can only be traversed in one direction. The stationary limit surface can be traversed in both directions, but in ergosphere region it is impossible to remain stationary. The singularity of the Kerr black hole is a ring singularity along the edge of $r = 0$.

direction. We can see this for example by looking at a photon that is emitted in the ϕ direction at a radius r in the equatorial plane $\theta = \pi/2$. The null condition there is

$$ds^2 = 0 = g_{tt}dt^2 + g_{t\phi}(dtd\phi + d\phi dt) + g_{\phi\phi}d\phi^2\,, \tag{3.170}$$

which implies that

$$\frac{d\phi}{dt} = -\frac{g_{t\phi}}{g_{\phi\phi}} \pm \sqrt{\left(\frac{g_{t\phi}}{g_{\phi\phi}}\right)^2 - \frac{g_{tt}}{g_{\phi\phi}}}\,. \tag{3.171}$$

On the stationary limit surface $g_{tt} = 0$ so that we have the two solutions

$$\frac{d\phi}{dt} = 0\,, \quad \frac{d\phi}{dt} = \frac{a}{2G^2M^2 + a^2}\,. \tag{3.172}$$

A photon emitted directly against the black hole rotation is thus able to just remain static, but a massive observer will necessarily rotate with the black hole. This is due to frame dragging, i.e. the rotation of the black hole ‘drags’ space-time along, fast enough so that nothing can resist. We saw frame dragging previously in the weak field limit for the gravitomagnetic effect (2.53).

Again as for RN, in between the outer and inner horizon t and r change roles, so that an observer has to move forward the inner horizon. Once past the inner horizon however, t and r change back to normal and the observer can move in principle freely in r again. It is then possible to avoid the singularity and to continue along a symmetric path that exits through new horizons to a new asymptotically flat region that is a copy of the original universe, as illustrated by the blue curve in Fig. 3.9.

The curvature singularity where the Kretschmann scalar is diverging is not at $r = 0$, but at $\rho = 0$. As $\rho^2 = r^2 + a^2\cos^2\theta$ we have that ρ vanishes only if

$$r = 0 \text{ and } \theta = \pi/2\,. \tag{3.173}$$

However, $r = 0$ is not actually a point in space but a disc, and $\rho = 0$ is the edge of the disc. The Kerr singularity is thus a ring singularity. It can be shown that it is possible to pass through the ring and to exit in another asymptotically flat space-time. However, that space-time has $r < 0$, which implies that Δ never vanishes and that thus there are no horizons (the ‘antiverse’ in the right-hand panel of Fig. 3.9). That region is different from our normal universe, and has other problematic features, like closed time-like curves that allow going back in time after going around along ϕ. This means that an observer could meet herself at an earlier time, which tends to be a bad idea.

But as for Schwarzschild, it is unlikely that all these strange features of Reissner-Nordström and Kerr black holes actually exist in black holes that are formed through stellar collapse (except for the ergosphere which lies outside the horizon). In addition, some of the interior features of the maximally extended RN and Kerr black holes are probably unstable to perturbations.

Black-hole thermodynamics

Thanks to the existence of a horizon, isolated classical black holes are indeed completely black and do not emit any radiation whatsoever. In the early 1970's a number of results were proven for classical black holes that looked surprisingly similar to the laws of thermodynamics:

- **0th Law:** The surface gravity κ must be uniform over the horizon of a black hole in stationary equilibrium.
- **1st Law:** $dM = \kappa/(8\pi) dA_{\rm BH} + \Omega dJ$
- **2nd Law:** $\Delta A_{\rm BH} \geq 0$

Here $A_{\rm BH}$ is the area of the event horizon, Ω the angular velocity and J the angular momentum, and one uses conventionally natural units where $G = c = \hbar = k_B = 1$ so that Newton's constant G is not present explicitly in the first term of the first law. The first law can contain additional work terms, for example ΦdQ if the black hole is charged, with Φ the electrostatic potential and Q the electric charge. The surface gravity is the acceleration at the surface of an object. Mathematically is defined for a static Killing horizon[9] with Killing vector $\mathbf{K}$ through

$$K^\mu \nabla_\mu K^\nu = \kappa K^\nu \,, \tag{3.174}$$

evaluated at the horizon. For a non-rotating non-charged (Schwarzschild) black hole of mass M we find $\kappa = GM/R_S^2 = c^4/(4GM)$.

The laws suggest that the surface gravity κ plays the role of a temperature and $A_{\rm BH}$ that of an entropy[10] In 1974 Hawking computed the evolution of quantum fields in the geometry of a forming black hole; he found that at late times, when the black hole had formed and was stationary, there was a steady outgoing blackbody radiation with temperature

$$T_{\rm BH} = \frac{\kappa}{2\pi} \frac{\hbar}{k_B c} \,. \tag{3.175}$$

A simplified explanation for the existence of Hawking radiation is the following: the vacuum in quantum theory is not a static object, but instead virtual pairs of particles and antiparticles are formed all the time. Near a horizon it is possible that one of the particles falls into the horizon while the other one escapes and in the process becomes a "real" particle. The black hole then has to pay the bill as energy is conserved.

In order for the first law of black hole thermodynamics to correspond to the first law of thermodynamics, if the temperature is given by (3.175) then the entropy has to be

$$S_{\rm BH} = \frac{A_{\rm BH}}{4} \frac{k_B c^3}{G\hbar} \,. \tag{3.176}$$

[9] A Killing horizon is the surface where a Killing vector field is null along a null hypersurface, we saw above that this is the case for the Schwarzschild event horizon, which is therefore also a Killing horizon.

[10] A heuristic argument why the entropy should be related to the surface of the horizon is the following: Let us try to build a black hole from photons. The photon entropy is $S_\gamma \sim U/T$, also given in (3.178). The number of photons scales the same way, $N_\gamma \sim U/E_\gamma \sim U/T$. The equilibrium state maximises the entropy. What is the maximal number of photons that we can squeeze inside a black hole with radius R_S? As we need $\lambda \lesssim R_S$ we find $S_{\rm max} \sim M_{\rm BH}/E_{\gamma,{\rm min}} \sim M_{\rm BH}\lambda_{\rm max} \sim M_{\rm BH} R_S \sim R_S^2 \sim A_{\rm BH}$.

In natural units the entropy is simply a quarter of the area of the black hole horizon. We notice that in the classical limit $\hbar \to 0$ (and keeping κ constant) the black hole temperature vanishes and the entropy becomes infinite. Black hole thermodynamics is therefore clearly a quantum effect. Not only black holes, but all event horizons have an associated temperature and entropy.

The status of the third law for black holes is less clear. The limit $T \to 0$ corresponds to $\kappa \to 0$. For a Schwarzschild black hole this means that also the entropy vanishes, and the *cosmic censorship* conjecture, which says that no "naked singularities" can exist, then implies that Schwarzschild black holes with vanishing surface gravity (and thus vanishing size of the horizon) cannot exist. However, extremal rotating or charged black holes (with maximal angular momentum or charge) also have vanishing surface gravity but not a vanishing entropy. Nonetheless there is some evidence that also for black holes the $T = 0$ limit cannot be reached in practice.

The black hole entropy can also be expressed in units of the Planck length $l_P = \sqrt{\hbar G/c^3} \approx 1.6 \times 10^{-35}$m:

$$S_{\mathrm{BH}} = k_B \frac{A_{\mathrm{BH}}}{4l_P^2}\,. \tag{3.177}$$

The entropy is thus given (up to the factor $1/4$) by the horizon area of the black hole measured in units of a "Planck area" l_P^2. The radius of the event horizon of a non-rotating black hole is given by the Schwarzschild radius $R_S = 2GM/c^2 \approx 3M/M_\odot$km. The entropy of an astrophysical black hole (i.e. black holes of the mass of typical stars) is therefore huge, $S_{\mathrm{BH}} \approx 10^{76}(M/M_\odot)^2$. From the observed motion of stars in the center of the Milky Way galaxy we can infer that an object of several million solar masses is located there, an object that is near certainly a black hole (as nothing else could be as dense and stable). Such a supermassive black hole has thus an entropy of the order of $S_{\mathrm{BH}} \approx 10^{88}$, and astronomical observations indicate that most, if not all, galaxies have such a black hole at their center.

From your thermodynamics lecture you may recall that the entropy of black body radiation is

$$S_r = \frac{4a_{\mathrm{SB}}}{3} V T^3 = \frac{4}{3}\frac{U(T)}{T}\,, \tag{3.178}$$

with the radiation constant $a_{\mathrm{SB}} = \pi^2 k_B^4/(15\hbar^3 c^3)$, and where $U(T)$ is the energy of the photon gas. Since the (average) energy of a photon is proportional to T, we see that the entropy of the background radiation is proportional the number of photons with a pre-factor of order unity (actually the total entropy of particles in the Universe is about 7 times the photon number). There are currently about 400 photons per cm^3 in the cosmic microwave background radiation. The radius of the observable universe (taken to be the Hubble radius c/H_0) is about 14×10^9 light years corresponding to a volume of about 10^{31} ly^3 or with 1 ly $\approx 10^{16}$m this is about 10^{85} cm^3. The entropy of the background radiation is thus comparable to the entropy of the black hole at the center of the Milky Way alone, and since there are billions of galaxies like our own, the total entropy of the Universe today is completely dominated by the entropy of the supermassive black holes.

Inserting the surface gravity and the units explicitly, the temperature of a Schwarzschild black hole is given by

$$T_{\mathrm{BH}} = \frac{\hbar c^3}{8\pi k_B G M} \approx 6 \times 10^{-8} \frac{M_\odot}{M} \mathrm{K}\,, \tag{3.179}$$

i.e. astrophysical black holes have temperatures of about 10^{-7} K, much lower than the temperature of the CMB of 2.7K. This means that such black holes are not actually shrinking by emitting Hawking radiation, but instead growing by absorbing the background radiation! Also, supermassive black holes not only contain most of the entropy of the Universe, they are also the coolest objects in the Universe.

How quickly would isolated black holes evaporate? A black body with a temperature T and a surface area A in empty space radiates with a power

$$\frac{dE}{dt} = \sigma_{\rm SB} T^4 A \tag{3.180}$$

with $\sigma_{\rm SB} = a_{\rm SB} c/4$, based on the Stefan-Boltzmann law. From $E = Mc^2$, implying that $dE = -dMc^2$, and using $A = A_{\rm BH} = 16\pi M^2$, $T = 1/(8\pi M)$ (switching to natural units from now on) this corresponds to a mass loss of

$$\frac{dM}{dt} = -\frac{a_{\rm SB}}{256\pi^3}\frac{1}{M^2} \propto M^{-2} \tag{3.181}$$

so the total evaporation time t_e is

$$t_e = \int_0^{t_e} dt \propto -\int_M^0 dM' M'^2 \propto M^3 \,. \tag{3.182}$$

The evaporation will therefore end in finite time (for black holes that emit more radiation than they absorb). The age of the Universe is about 10^{10} years, which places an upper bound of mass of black holes that could have evaporated by today. That bound is $M \approx 10^{12}$kg, much less than the mass of the sun of about 2×10^{30}kg – it is about 1 km^3 of rock or water. Since astrophysical processes generate black holes of some reasonable fraction of a solar mass, only black holes that were formed by a different process (e.g. in the early universe, so called *primordial black holes*) can have evaporated by today.

Towards the end of the evaporation process the black hole mass approaches the Planck mass. At this point the analysis above is no longer valid as the back-reaction of the radiation on the black hole as well as quantum effects will become important. What actually happens during the final stages of black hole evaporation is not known.

The specific heat of a black hole is negative:

$$C = \frac{dE}{dT} = \frac{dM}{dT_{\rm BH}} = -8\pi M^2 < 0 \,. \tag{3.183}$$

This means that a black hole is thermodynamically unstable: If the environment of the black hole is cooler than $T_{\rm BH}$ then the black hole will lose energy and heat up further. In the opposite case it absorbs energy from the environment and becomes cooler, radiating less. In this way, the black hole moves further away from equilibrium. This is typical for gravitating systems (and necessary to allow structure to form from an initial thermal and almost homogeneous state). There are some ways to create thermodynamically stable black holes, for example by adding a charge.

During an evaporation process, the black hole is shrinking and so appears to be violating the second law. But we have of course to consider the total system, in this case the black hole and the emitted radiation. (This is also the case when matter collapses to form a black hole, but we have already seen that the entropy of the black hole is much larger than the entropy of the matter that creates it.) The radiation emitted by the black hole at temperature T in a short time interval Δt is

$$\Delta E = A_{\rm BH} a_{\rm SB} T_{\rm BH}^4 \Delta t \tag{3.184}$$

and from (3.178) the corresponding entropy increase in the radiation field is

$$\Delta S_r = \frac{4}{3} A_{\rm BH} a_{\rm SB} T_{\rm BH}^3 \Delta t = \frac{4}{3}\frac{\Delta E}{T_{\rm BH}} \,. \tag{3.185}$$

The change of black hole entropy on the other hand is

$$\Delta S_{\rm BH} = \frac{1}{4} dA_{\rm BH} = 8\pi M \Delta M = -8\pi M \Delta E = -\frac{\Delta E}{T_{\rm BH}} \,. \tag{3.186}$$

We therefore have that $\Delta S_r/|\Delta S_{\rm BH}| \simeq 4/3 > 1$, i.e. the entropy in the radiation emitted by the black hole is larger than the entropy lost by the black hole by a factor of $4/3$. The second law is therefore safe.

Black holes are interesting because they appear to follow thermodynamic rules, but we do not know their microscopic description, in a statistical / quantum mechanics sense. The entropy may be linked to the number of microstates inside the black hole (as would be expected from statistical mechanics) or maybe to the information loss due to the presence of a horizon, i.e. to information entropy. The latter notion could be related to entanglement entropy due to the fact that our description of quantum objects is incomplete when horizons are present and that we have to integrate out the unknown degrees of freedom inside the horizon. In general the entanglement entropy is different from the entropy due to the possible microstates. However, for black holes the two may coincide.

The loss of information when matter forms a black hole is an especially interesting question: The no-hair theorem says that the most basic uncharged black hole is described only by its mass and angular momentum. However, the matter that falls into the black hole can be in an arbitrarily complicated configuration. If the black hole evaporates completely by emitting purely thermal uncorrelated radiation, then all this information about the initial state appears lost. This contradicts a cherished notion from quantum mechanics, namely that the evolution of quantum systems is always unitary (i.e. conserves probability). While a number of proposals have been put forward on how to explain away the problem of information loss, none of them have really proven satisfactory. However, in an interesting twist a link has been found recently between specific black hole solutions (that are not really relevant for our universe, but nonetheless exhibit in principle the same information loss problem) and the explicitly unitary evolution of a quantum theory, the so-called AdS/CFT correspondence. Unfortunately, although this would seem to indicate that black holes do preserve information (at least in some cases), the AdS/CFT correspondence does not shed any light on how this happens. The verdict on whether information is lost in black holes is therefore still open.

Bibliography

[1] Robert M. Wald, *General Relativity*, Chicago University Press 1984.

[2] Norbert Straumann, *General Relativity*, 2nd Edition, Graduate Texts in Physics, Springer Verlag, Berlin 2013.

[3] Ruth Durrer, *General Relativity*, http://fiteoweb.unige.ch/˜durrer/courses/relaE.pdf.

[4] Sean Carroll, *Spacetime and Geometry: An Introduction to General Relativity*, Pearson 2014.

[5] D. Lovelock, J. Math. Phys. **12**, 874 (1972).

[6] John David Jackson, *Classical Electrodynamics*, Second Edition, Wiley 1975.

[7] C. M. Will, Living Rev. Rel. **17** (2014) 4 doi:10.12942/lrr-2014-4 [arXiv:1403.7377 [gr-qc]].