Advanced General Relativity and Quantum Gravity

Lecture in SS2024 at the KFU Graz

Axel Maas

Contents

1 Introduction **1**

2 General relativity **3**

2.1 Kinematics . . . 3

2.1.1 Manifold structure . . . 3

2.1.2 Parallel transport . . . 7

2.1.3 Distances . . . 10

2.1.4 Diffeomorphism symmetry . . . 13

2.1.5 Geodesic distances . . . 15

2.2 Dynamics in pure gravity . . . 16

2.3 Dynamics in the presence of matter . . . 18

2.4 The vielbein formalism and torsion . . . 19

2.5 Lagrangian formulation . . . 23

2.6 Energy, Fourier-space and physical observables . . . 27

3 Special solutions **29**

3.1 The initial value problem . . . 29

3.2 Identifying gravity . . . 34

3.3 Maximally symmetric solutions . . . 36

3.4 Big-bang solutions . . . 39

3.5 Schwarzschild black hole and Kruskal coordinates . . . 43

3.6 Kerr black hole . . . 43

3.7 Gravitational waves . . . 43

3.8 Wormholes, geons and other exotic solutions . . . 43

3.9 ADM construction and momenta . . . 43

4 Canonical quantum gravity **44**
4.1 Perturbation theory and its failure 44
4.1.1 Perturbative quantization 44
4.1.2 One-loop 44
4.1.3 Physical observables and augmenting perturbation theory 44
4.2 Beyond the Einstein-Hilbert Lagrangian 44
4.3 Functional renormalization group 44
4.4 Dynamical triangulation 44
4.5 Gravity as the square of Yang-Mills theory 44

5 Beyond canonical quantum gravity **45**
5.1 Loop-quantum gravity 45
5.2 Causal sets 45
5.3 Non-commutative geometry 45
5.4 String theory 45

Chapter 1

Introduction

General relativity is often considered to be quite enigmatic, due to its separation from quantum physics. It is also often mystified due to it being linked to everyday experience like time, as it changes these. The aim of this lecture is to provide a link between quantum theories and general relativity, and at the same time demystify it.

For that purpose, the first part of this lecture will provide a more in-depth introduction to classical general relativity. Especially, it will highlight how general relativity, despite its appearance, is really a theory very similar to other gauge theories like Yang-Mills theories. What makes it look apart is that we are very much caught in a very special solution, and have not the possibility to play around with initial conditions, as in electrodynamics. To this end, special solutions, like black holes and the universe, will be discussed in some detail, to highlight what are physical concepts and what are just auxiliary constructions.

The other challenging question is how to quantize gravity. If one prefers to stick with the idea that quantum field theory is essentially the correct approach to nature, this becomes inevitable. This will be assumed here. What makes this an assumption, and quantum gravity somewhat more shaky than quantum field theory, are two aspects. The one is the total lack of experimental insights about what could be the nature of quantum gravity. This is not a problem of principle, but rather a consequence of the smallness of the parameters of classical general relativity. While it is unclear when this technical barrier can be surpassed, one of the currently best guesses are gravitational waves. They will therefore serve an an intermediary.

The other problem is that, while formally insufficient, standard perturbative quantization of even the simplest quantum gravity theory fails. The reason is the lack of perturbative renormalizability. Furthermore, many decisive concepts of quantum field theory, like four-momentum, requires a very different approach. This made quantization a challenge, as it either requires to perform some kind of non-perturbative quantization or switch to a different approach to quantization. Only within the last few decades, non-perturbative approaches started to yield convincing results. On the other hand, other approaches like loop quantum gravity and string theory, made also progress. But again, without any experimental insight and the classical limit being fairly generally reproducible, no decision is yet possible. Therefore, this lecture aims at giving a brief overview of the various possibilities.

As a consequence, there is a vast amount of literature available on the subject. As usual, the suitability is highly personal. For the sake of completeness, the following books and articles have been used in the preparation of this lecture:

- Ambjorn et al., “Nonperturbative Quantum Gravity”, Phys. Rept. 519 p127 (2012)
- A. Asthekar et al., “Loop quantum cosmology: A status report”, Class. Quant. Grav. 28 p213001 (2011)
- Freedman et al., “Supergravity”, Cambridge
- Hehl et al., “General relativity with spin and torsion”, Rev. Mod. Phys. 48, p393 (1976)
- Misner et al., “Gravitation”, Freeman
- J. Polchinski, “String Theory”, Cambridge
- Straumann, “General relativity”, Springer

Chapter 2

General relativity

2.1 Kinematics

2.1.1 Manifold structure

As every theory, also general relativity can be split into a kinematical part, how the physical system is described, and a dynamical part, how things happen. The main difference is that the kinematical part is not embedded in an arena, e. g. the space-time of quantum field theory, but exists without any embedding. This is also one of the major conceptual challenges in understanding general relativity. There is nothing left to stand on and prepare the system. But this is not entirely true, as will be seen. However, what remains true is that general relativity does not allow for the notion of an outside of the system.

The kinematic structure of general relativity is a topological pseudo-Riemannian manifold $\mathbb{M}$. As a manifold, any element of the manifold m is in one-to-one correspondence to an element x of (a patch of) $\mathbb{R}^d$, $m(x) = x^{-1}(m)$ and $x(m) = m^{-1}(x)$ both exist and are well defined. The dimension d is the minimum required to allow for such a one-to-one correspondence. These elements and also their identifiers x will be both called events in this lecture. Moreover, the manifold structure requires that there exists a local isomorphism between the manifold and the $\mathbb{R}^d$ in the sense that overlapping subsets in the manifold are mapped to overlapping subsets in $\mathbb{R}^d$, where the overlap is identical, i. e. contains the same sets of m and x such that $m(x)$ holds for all elements in the overlap. This allows to make a statement about neighborhood relations, though there is not yet any notion of distance. An example of the concept is the surface of a sphere, where points on the surface are the elements of the manifold, and the angular coordinates θ and ϕ form a patch in $\mathbb{R}^2$ on which overlapping subsets are mapped. Note that due to the south pole and the north pole this is not a square patch, but sometimes single points are added along a line.

It is useful to introduce coordinates $X(m) = X(m(x)) = X(x)$ on a manifold. In particular, they can be chosen, e. g., to realize the open set structure in a coordinate language. E. g., for the sphere, coordinates could be chosen to be latitude and longitude. In section 2.2 it will be required that the manifold always allows to at least locally introduce a coordinate system, which is the same as that of Minkowski space-time, i. e. the manifold being locally Lorentzian.

Because only the overlapping set relation needs to be maintained, and the structure of the patch in $\mathbb{R}^d$ can be involved, it is in general not possible to express the coordinates on a manifold globally. To cover it, rather multiple coordinate systems are required, which overlap in both the manifold and the underlying $\mathbb{R}^d$. They are then related by transfer functions, which map the coordinates into each other, i. e. $X_2(X_1) = X_2(X_1^{-1}(x)) = X_2(x)$.

In the sphere case, this can e. g. take the following form. Latitude and longitude are ill-defined at either pole. So, a second coordinate system is needed. Labelling the elements of the manifold set by their coordinates in three dimensions, and the underlying $\mathbb{R}^2$ path by the angle $\phi \in [0, 2\pi)$ and $\theta \in [-\pi/5, 4\pi/5)$, the coordinate systems are given by

$$\begin{aligned}
\vec{r} &= \begin{pmatrix} \cos\phi_1 \sin\theta_1 \\ \sin\phi_1 \sin\theta_1 \\ \cos\theta_1 \end{pmatrix} \\
\vec{t} &= \begin{pmatrix} \cos(\theta + 2\pi/5) \\ \sin(\phi)\sin(\theta + 2\pi/5) \\ \cos(\phi)\sin(\theta + 2\pi/5) \end{pmatrix}.
\end{aligned}$$

where the first coordinate system lives in the θ strip $[\pi/5, 4\pi/5)$, and the second one in $[-\pi/5, 2\pi/5)$. The overlap exist for the strip $(\pi/5, 2\pi/5)$. In this case, the transfer function is

$$\vec{t}(\vec{r}) = \begin{pmatrix} \cos\left(\cos^{-1}(r_3) + \frac{2\pi}{5}\right) \\ \sin\left(\tan^{-1}\frac{r_1}{r_2}\right)\sin\left(\cos^{-1}(r_3) + \frac{2\pi}{5}\right) \\ \cos\sin\left(\tan^{-1}\frac{r_1}{r_2}\right)\sin\left(\cos^{-1}(r_3) + \frac{2\pi}{5}\right) \end{pmatrix}. \tag{2.1}$$

In the relevant strip both the cosine and the tangens are single-valued. This transfer function between $\vec{r}_2$ and $\vec{r}_1$ is thus invertible and differentiable in the overlap. Thus, it is also possible to express $\vec{r}_2$ as a function of $\vec{r}_1$, and likewise, both ϕ and θ can be uniquely extracted, if so desired.

The minimal set of coordinate system to cover the whole manifold is called an atlas. Note that a coordinate system also offers a reparametrization, e. g. by a rotation. This freedom is independent and applies to each coordinate system separately. Finally, for the example the manifold has been embedded[1] in a higher-dimensional space. This made it easier to express the difference between the coordinates r_i and the underlying $\mathbb{R}^d$. In fact, in such an embedding a three-dimensional coordinate system could be introduced, x, y, and z, which would cover the whole sphere. However, it can be proven that this not possible in general. Hence, it is necessary to stick with transfer functions.

If the transfer functions are, as in the example, differentiable as a function of the parameters, the manifold is called differentiable. In general relativity the manifold will always be required to be differentiable.

The manifold is assumed to be topological, i. e., there exists as distance measure $d(m_1, m_2) = d(m_1(x_1), m(x_2)) \to \mathbb{R}$, which allows to determine a coordinate-independent distance between two elements of the manifold, and thus two events. This is called usually the invariant length element ds, and can be calculated in terms of coordinates, $ds = d(X(x), Y(y))$. It is not assumed that the distance measure is positive-definite, i. e. ds can have either signs. It may also be zero even for any pairings of events. However, it is required that $d(m, m) = 0$. Such a manifold is called pseudo-Riemannian. If $d \geq 0$ would be satisfied, it is called Riemannian.

What will be required axiomatically for general relativity, however, is that the distance measure is locally Minkowski, implying the local existence of a metric. This restricts the possible topological manifolds, as this requires the distance measure to be locally a linear form. This implies that for any pair of events x and y, there exists a δ

[1] A two-dimensional coordinate system would be, e. g., latitude and longitude on the some broad ring, and then shift the meridian for the other broad ring, such that in neither case the north-pole and south-pole of the respective meridian would be part of the covered sphere. However, to avoid confusing both angular systems, it is more useful for now to use the embedding, such that already the dimensionality distinguishes the underlying $\mathbb{R}^2$ and the coordinate system.

such that if $|x-y|<\delta$ in the Euclidean norm[2] then

$$\begin{aligned} d(X(x),Y(y)) &= (X_\mu - Y_\mu)\eta^{\mu\nu}(X_\nu - Y_\nu) \\ X &= x \\ Y &= y \end{aligned} \tag{2.2}$$

where the coordinates are locally introduced trivially and $\eta_{\mu\nu}$ is the usually Minkowski metric of special relativity. Thus, for any fixed coordinate system, this generalizes to the existence of a metric $g_{\mu\nu}(X)$

$$d(X,Y) = (X_\mu - Y_\mu)g^{\mu\nu}(X)(X_\nu - Y_\nu),$$

but $d(X,Y)$ independent of the choice of coordinate system. It is required here that $g_{\mu\nu}(X) = g_{\mu\nu}(Y)$. Thus distances are the same no matter from where it is started to measure. As will be seen in section (2.1.3), this further restricts the possible manifolds.

Such a metric is called compatible. The distances $d(X,Y)$ are called invariant distances. In addition, for the present lecture, it will be defined that $d(X,Y)<0$ is called space-like, $d(X,Y)>0$ is called time-like, and $d(X-Y)=0$ is called light-like. That requires $\eta = \mathrm{diag}(1,-1,-1,-1)$. The local isomorphism always allows to select a patch to introduce a coordinate system without transfer functions in (2.2). This is no longer true if the distance is not infinitesimal. This will be discussed in section 2.1.3.

To give an example, consider again the sphere. Because the sphere is not a pseudo-Riemannian manifold but a Riemannian manifold, the local metric will be the Euclidean metric $\delta_{\mu\nu}$ instead. The distance measure will be chosen such that the distance between two elements of the manifold will be the line distance on the sphere. Thus, going the sphere around once on a grand circle, which will only be looked at in section 2.1.5, will yield a distance of 2π. This requires the usual length element

$$ds^2 = d\theta^2 + \sin^2\theta d\phi^2. \tag{2.3}$$

Consider first the first coordinate system. Because of the embedding, the necessary metric is

$$g^r(\theta,\phi) = \frac{1}{\cos^2\theta}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -\frac{\sin^4\theta}{\sin^2\phi} \end{pmatrix} \tag{2.4}$$

[2]By construction, there is always a sufficiently small patch in which the elements of the manifolds are isomorphic to the patch in the underlying Euclidean space.

which can be expressed in terms of the coordinates as

$$g^r(\vec{r}) = \frac{r_3^2}{r_1^2 + r_2^2 + r_3^2} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -\frac{(1-r_3^2)^2(r_1^2+r_2^2)}{r_2^2(r_1^2+r_2^2+r_3^2)^2} \end{pmatrix},$$

showing that the singularity is located along the 2-axis. Likewise, using the coordinate system $\vec{t}$ a metric is obtained where the angle θ would be shifted. Since in the line element (2.3) must remain the same in terms of the underlying $\mathbb{R}^2$, this yields

$$\begin{aligned} g_{11}^t = g_{22}^t &= \frac{\sin^2\phi - \cos^2\theta\sin^2\theta\tan^2\left(\theta - \frac{\pi}{10}\right)}{\sin^2\left(\theta - \frac{\pi}{10}\right)\cos^2\phi - \cos^2(2\phi)} \\ g_{33}^t &= \frac{-\cot^2\phi + \sin^2\theta\left(\csc^2\phi + \tan^2\left(\theta - \frac{\pi}{10}\right)\right)}{\sin^2\left(\theta - \frac{\pi}{10}\right) - \cos^2\phi\sin^2\left(\theta - \frac{\pi}{10}\right)\tan^2\left(\theta - \frac{\pi}{10}\right)\left(\cot^4\phi - 1\right)}, \end{aligned}$$

and all other entries vanishing. This can again be expressed in terms of the coordinates, yielding an entirely coordinate-dependent expression. In particular, in terms of coordinates the invariant distance measure (2.3) differ, but always yield the same result. E. g. in the coordinate system r, the line element take the form

$$ds^2 = r_3^2(dr_1^2 + dr_2^2) - \frac{(r_1^2 + r_2^2)r_3^2(r_3^2 - 1)^2}{r_2^2(r_1^2 + r_2^2 + r_3^2)} dr_3^2 \tag{2.5}$$

and can be calculated now entirely in terms of the coordinates.

In a general manifold, (2.2) will only be possible in an infinitesimally small neighborhood. To extent the concept of distance requires the introduction of paths.

2.1.2 Parallel transport

An important concept is that of a path. A path is a continous sequence of events labeled by a real number τ, $m(\tau)$ or equivalently $x(\tau)$. This creates a path in the manifold $X(x(\tau)) = X(\tau)$. It may be necessary to switch the coordinate system along any finite path. However, for any path there will be a patch around $\tau = \tau_0$ in which a single coordinate system is sufficient. This allows to introduce the derivative of the path in the manifold as $\partial_\tau X(\tau)$, as here the difference of coordinates make sense. This defines a direction, and consequently can be used to attach locally a vector space with vectors X_μ, which is spanned by dimension linearly independent paths, a tangent space. If the number of dimensions of the tangent space is not everywhere the same and equal to the number of dimensions of the manifolds, this is called the a degeneracy or singularity of the manifold. Only at events like black holes such things may occur in general relativity.

Such tangent vectors will change under the change of the coordinate system by a transformation Λ to ΛX. Vectors which change in this way will be called covariant vectors and labeled by an upper index. Quantities, which change by the inverse Λ^{-1} will be called contravariant vectors, and labeled by a lower index.

However, the vectors are now defined in terms of a differential at a single point. Two vectors defined in the tangent space at different events will not be comparable, as they formally belong to different space. To compare vectors V at two different events requires to parallel transport them, taking into account the change of tangent space. While mathematically this remains a space of fixed dimensionality, the base vectors change by the parallel transport along a path, due to the change of directional differentials. Thus, they will only remain the same if these derivatives do not change. This is not the same as an arbitrary base transformation, as it depends on the path taken. Thus, what is necessary is to track how the base vectors change along the path.

Moving an infinitesimal distance, this can at most be an infinitesimal, and thus linear change, in the vector. The most general possibility is quantified by the affine connection Γ,

$$dV^\mu = \Gamma^\mu_{\nu\rho}(X)V^\nu dX^\rho. \tag{2.6}$$

To yield full sense, this equation needs to be divided by $d\tau$, and thus describes how a vector changes, when transported along the path $X(\tau)$. If the boundary of a coordinate system is traversed, this needs to be taken into account. However, it is always possible to first switch to a coordinate system, which covers both infinitesimally separate points. Thus, when traversing some distances which requires a change of coordinate system, the best choice is to move in one coordinate system until entering the overlap region, then stop and switch to the new coordinate system, and then move on in the second coordinate system. This avoids the need to express differentials across boundaries of coordinate systems, and is always possible due to the required overlap.

The affine connection Γ is a feature of the manifold. The path and coordinate system only enters in terms of dX^μ, $d\tau$, and dV^μ is determined by the application. Thus (2.6) cleanly separates the manifold, the coordinate system, and the object. Of course, as tensors of higher rank can be introduced into the tangent space, (2.6) can be generalized to tensors of arbitrary rank. This is done using the corresponding tensor product rules.

Note that the affine connection is, currently, independent from the topology, and thus not related to the metric. It is a measure of how the manifold bends along a path. The antisymmetric part of Γ

$$S^{\mu}_{\nu\rho} = \frac{1}{2}\left(\Gamma^{\mu}_{\nu\rho} - \Gamma^{\mu}_{\rho\nu}\right) \tag{2.7}$$

is Cartan's torsion tensor. Whether S is zero or not has far-reaching consequences, and is a property of the manifold in question. In general relativity proper, it is required to be zero. As

$$S^{\mu}_{\nu\rho} V^{\nu} dX^{\rho} = \frac{1}{2}\left(dV^{\mu} - \Gamma^{\mu}_{\rho\nu} V^{\nu} dX^{\rho}\right)$$

this shows that it describes the vorticity or torsion, of the manifold. This expression vanishes, if for the change dV^{μ} it does not matter whether a left-screw or a right-screw change is made.

Since (2.6) allows to quantify the change of a vector in a given direction, is thus allows to define a covariant derivative on vectors

$$D^{\mu} V_{\nu} = \left(\partial_{\mu}\delta^{\lambda}_{\nu} - \Gamma^{\lambda}_{\mu\nu}\right) V_{\lambda}, \tag{2.8}$$

For a scalar quantity, this reduces to

$$D^{\mu}\phi = \partial^{\mu}\phi,$$

as only the argument needs to be transported, but scalars have no directional properties, and are thus not affected. More general tensors, which are constructed by a tensor product $X_{\mu\nu} = V_{\mu} W_{\nu}$, this yields

$$\begin{aligned} D^{\mu} X_{\nu\rho} &= V_{\nu} D^{\mu} W_{\rho} + \left(D^{\mu} V_{\nu}\right) W_{\rho} = \partial^{\mu} X_{\nu\rho} - V_{\nu}\Gamma^{\lambda}_{\mu\rho} W_{\lambda} - \Gamma^{\lambda}_{\mu\nu} V_{\lambda} W_{\rho} \\ &= \partial^{\mu} X_{\nu\rho} - \Gamma^{\lambda}_{\mu\rho} X_{\nu\lambda} - \Gamma^{\lambda}_{\mu\nu} X_{\lambda\rho}. \end{aligned}$$

and likewise for higher tensors. This defines the action of D_{μ} on tensors of arbitrary rank.

It is not necessary that when moving a vector along a closed path, it remains unchanged. This information is encoded in whether covariant derivatives can be exchanged for a differentiable function. On a flat manifold, i. e. one on which $\Gamma = 0$, any twice differentiable function will behave as $(D_{\mu}D_{\nu} - D_{\nu}D_{\mu})V = 0$. This is encoded in the Riemann curvature tensor defined as

$$\begin{aligned} \left[D_{\mu}, D_{\nu}\right] V^{\lambda} &= R^{\lambda}_{\rho\mu\nu} V^{\rho} \\ R^{\lambda}_{\rho\mu\nu} &= \partial_{\mu}\Gamma^{\lambda}_{\nu\rho} - \partial_{\nu}\Gamma^{\lambda}_{\nu\rho} + \Gamma^{\lambda}_{\mu\sigma}\Gamma^{\sigma}_{\nu\rho} - \Gamma^{\lambda}_{\nu\rho}\Gamma^{\sigma}_{\mu\rho}. \end{aligned} \tag{2.9}$$

It therefore characterizes the manifold. As the Riemann tensor is given entirely in terms of the affine connection, it does not create another independent characterization of the manifold, but it is a combination often useful.

It should be noted that at this point, these are general statements. As the manifold is not specified, it is not yet possible to calculate any of these quantities. Before switching to examples, it is, however, useful to first consider another concept, distances.

2.1.3 Distances

Distances were already introduced as a local concept within an infinitesimal ball. Combining them with parallel transport, it is possible to extend the concept. Given (2.6), an infinitesimal distance can be defined as

$$ds^2 = g_{\mu\nu}(X)dX^\mu dX^\nu, \tag{2.10}$$

where dX^μ are the infinitesimal distances between X and a neighboring element of the manifold in direction μ. The transformation properties under coordinate changes implies that there exist an inverse metric, based on the contravariant vectors, satisfying

$$g^{\mu\nu}(X)g_{\nu\rho}(X) = \delta^\mu_\rho, \tag{2.11}$$

locally. The condition (2.2) implies that there exists always a coordinate transformation that at some specified point X $g_{\mu\nu}(X) = \eta_{\mu\nu}$ and $g^{\mu\nu}(X) = \eta^{\mu\nu}$. However, in general at other points $X' \neq X$ the metric will not be Minkowski. Because coordinate transformations are required to be invertible, the sign of the determinant is coordinate-independent. This implies that $\operatorname{sgn}\det g = \operatorname{sgn}\det\eta$. However, a general coordinate transformation can include a rescaling, and thus $\det g$ is not invariant.

Due to the necessity of the metric to transform under coordinate transformations appropriately to maintain (2.11) invariant, this implies that the metric can be used to raise and lower indices as

$$\begin{aligned} g_{\mu\nu}V^\nu &= V_\mu \\ g^{\mu\nu}V_\nu &= V^\mu. \end{aligned}$$

This holds true for any quantity defined using (2.6), including the affine connection and the Riemann curvature tensor.

For a general manifold, the covariant derivative of the metric will not be zero,

$$Q_{\mu\nu\rho} = D_\mu g_{\nu\rho}.$$

Classical general relativity requires this tensor of non-metricity Q to vanish. The metric is then covariantly constant. If this would not be the case, measuring distances would depend on the point of reference from where to measure. General relativity is based on the experimental result that this is not the case, and thus requires the tensor of non-metricity to vanish as an empirical requirement. This restricts the class of possible manifolds. Thus, the direction of traversing a path will not change the invariant distance.

Now, consider finite distances. There is no a-priori reason why a finite distance will in general not be dependent of the path taken. The total distance traversed can be obtained by moving from point to point along a path $X(\tau)$, this yields that the total distance traveled along the path is given by

$$s = \int ds = \int\limits_{\tau_0}^{\tau_1} \frac{ds(\tau)}{d\tau} d\tau = \int\limits_{\tau_0}^{\tau_1} d\tau \sqrt{g_{\mu\nu}(X(\tau)) \frac{dX^\mu(\tau)}{d\tau} \frac{dX^\nu(\tau)}{d\tau}}, \tag{2.12}$$

which generalizes the eigentime. It is invariant under reparametrizations of the parameter τ. To show this invariance, take an arbitrary (but invertible) reparametrization $\tau' = \tau'(\tau)$. This implies

$$\begin{aligned} \dot{\tau}' &= \frac{d\tau'}{d\tau} \\ d\tau &= \frac{d\tau'}{\dot{\tau}'}, \end{aligned}$$

yielding the transformation property of the integral measure. For the derivatives follows then

$$\dot{X}^{\mu'}(\tau') = \dot{X}^{\mu}(\tau) \frac{d\tau}{d\tau'} = \dot{X}^{\mu} \frac{1}{\dot{\tau}'}$$

Hence the scalar product changes as

$$\dot{X}^{\mu\prime} \dot{X}_{\mu\prime} = \frac{1}{\dot{\tau}'^2} \dot{X}^{\mu} \dot{X}_{\mu}.$$

One power of $\dot{\tau}'$ is removed by the square root, and the remaining one is then compensated by the integral measure. This implies that without the square root the result would not be independent. As usual, this requires that also the integral limits $\tau_{0,1}$ are transformed.

The expression (2.12) is a path-dependent distance. To better understand the implication of this, consider the case of Minkowski metric, and various paths. Start with the path

$$r = (\tau, 0, 0, 0)^T$$

and always $\tau_0 = 0$ and $\tau_1 = 1$. This path is a time-like vector for every value of τ. This yields $s = 1$. If instead the path

$$r = (\tau^2, 0, 0, 0)^T$$

is chosen, but with the same τ_i and thus not being a reparametrization, the result is 3/4. Because the path is traversed at a different rate, the total elapsed eigentime, and thus time-like distance, changes[3].

Consider a path, which is light-like at every value of τ,

$$r = (\tau, 0, 0, \tau)^T.$$

The distance is zero, as the integrand is zero. The distance along a light-like path always vanishes. Again, this can be understood as a light-like vector can never be transformed into a rest frame, and thus no (eigen)time elapses for an object traveling along a light-like trajectory.

Something odd happens for a space-like path,

$$r = (0, 0, 0, \tau)^T.$$

Formally the expression (2.12) becomes imaginary. The reason is that when taking the root an assumption was made on the sign of the invariant distance. A meaningful result is obtained by using

$$s = -\int_{\tau_0}^{\tau_1} d\tau \sqrt{-g_{\mu\nu}\frac{dX^\mu(\tau)}{d\tau}\frac{dX^\nu(\tau)}{d\tau}},$$

taking instead. This yields the expected space-like distance -1 for the path.

What happens if a mixed path is chosen, which is partially time-like and partial space-like (or light-like)? In that case the argument of the root changes sign, leading to a seemingly complex distance. That can be partially avoided by stitching the total distance together from piece-wise definite behavior. While this formally possible, this is actually a conceptually quite non-trivial issue. A pure time-like path is physically sensible, as it determines the causal evolution along a world-line. A purely space-like path answers the question of distance of an observer at fixed eigenzeit. A stitched path will first move along a space-like direction and then hitch onto a worldline, and that possibly multiple times. There is therefore rarely a physics reason to consider this issue.

[3]This is also the resolution to the so-called twin paradox.

Switching back to the general case, this implies that the character of the distance is not only influenced by the path, but also how the metric changes along the line. However, the character of the distance can also be fixed when choosing locally coordinates such that the metric becomes the Minkowski metric. It does not appears sensible to change the nature of the distance, time-like, space-like or light-like, because of this. This leads to the concept of diffeomorphism symmetry.

2.1.4 Diffeomorphism symmetry

It is worthwhile to understand better what is actually meant when talking about coordinate transformations in the present case. Coordinate transformations on the manifold can be arbitrary, $X'(x) = X'(X(x))$, but needs to be invertible, $X(x) = X(X'(x))$. Since the manifold is differentiable, these transformations need to be differentiable themselves, though they may become involved due to the need for transfer functions. Distances, and thus the topological manifold structure, are required to remain invariant. This symmetry is hence called a diffeomorphism symmetry.

Now, any such transformation can necessarily be written as

$$X'(x) = X'(X(x)) = X(x) + \epsilon(X(x)), \tag{2.13}$$

and all changes are encoded in the event-dependent function ϵ. Likewise, the inverse transformation is then given by $-\epsilon(x)$. This also implies

$$\frac{\partial X'}{\partial X} = 1 + \frac{\partial \epsilon}{\partial X},$$

and thus the derivatives of ϵ define the deviation from no transformation.

The important insight is, however, that (2.13) is a local translation. Thus, general coordinate transformation is really a gauging of the global translation symmetry. It appears as if the Lorentz symmetry of the Poincaré group has simply vanished. This appears very strange, especially as the Lorentz group plays an important role for the spin of particles. The latter issue will be postponed until section 2.4.

As for the other, it should be noted that $\epsilon(X)$ can be arbitrarily split as

$$\epsilon_\mu(X) = \Lambda_{\mu\nu}(X)X^\nu + \delta(X). \tag{2.14}$$

Thus, any local translation can always be decomposed into a local rotation and a local translation, but not uniquely. However, this implies that the global Poincaré group is a subgroup of the local coordinate transformations.

However, the idea should not be that special relativity emerges as the limit of (2.14). As general relativity is a classical theory, the structure of the manifold is fixed as function of the initial conditions, in a certain sense, see section 3.1. Thus, special relativity should rather be interpreted as the case where the manifold determined by the initial conditions has a trivial atlas and has invariant distances such that $g = \eta$ can be introduced globally[4]. Because the Minkowski metric is invariant under Poincaré transformations, this symmetry emerges then as a symmetry of the obtained manifold. Thus, the structure of special relativity is a consequence of the dynamics to be introduced in section 2.2. This makes the situation quite different from the classical mechanics limit of special relativity, which is purely kinematical and not affected by the dynamics.

It that sense, diffeomorphism symmetry can be considered to be the gauge symmetry of translations. The gauge-invariant observables, analogues to the electric field and magnetic field of classical electrodynamics, are then the invariant distances, i. e. the topology of the manifold. As will be seen when introducing the dynamics, the metric can be given the role of the gauge field, though it is a rank two tensor now, rather than a rank one tensor in electrodynamics.

Note that this implies that the metric cannot be invariant under a diffeomorphism symmetry. In fact, since distances need (2.10) need to be invariant, it follows that

$$ds^2 = g_{\mu\nu} dX^\mu dX^\nu \overset{!}{=} g'_{\mu\nu} dX^{\mu'} dX^{\nu'}.$$

The coordinate differentials transform as

$$dX'^\mu = \Lambda^\mu_\nu dX^\nu = \frac{\partial X^\mu}{\partial X^\nu} dX^\nu = \left(\delta^\mu_\nu + \frac{\partial \epsilon^\mu}{\partial X^\nu}\right) dX^\nu$$

requiring

$$g'_{\mu\nu} = (\Lambda^{-1})^\alpha_\mu (\Lambda^{-1})^\beta_\nu g_{\alpha\beta} \tag{2.15}$$

and likewise its inverse. This is reminiscent of the situation in special relativity. The only difference is that Λ is not a Lorentz transformation, but a diffeomorphism, and that this is a local statement. It is, however, often more convenient to let g transform as Λ and dX^μ with Λ^{-1}, which is a mere shift of definition of Λ.

This resolves also the observation from the end of section 2.1.3. The character of the invariant distances traversed by a path remains invariant under a diffeomorphism symmetry. It is thus always possible to attribute to a path, or at least to parts of path, uniquely whether it is time-like or space-like. It is thus a feature of the path, not the

[4] Of course, in reality this will only be approximately true.

metric, just like in special relativity. The metric, in turn, is dictated by the manifold, and describes how 'fast' paths can be traversed, and thus how much eigentime expires, e. g., when moving from one place to another.

2.1.5 Geodesic distances

So far, distances have been introduced as a path-dependent concept. In fact, the non-vanishing Riemann tensor (2.9) shows that the concept of distance can no longer be path independent on an arbitrary manifold. Thus, the distance between two-events, no matter when it is time-like or space-like, is no longer uniquely defined. In particular, the (absolute) value can often be made arbitrarily large. E. g., on a sphere dense full orbits can be performed on a path between two elements, and therefore the distance can be made arbitrarily large.

It is, however, possible to pose a different question for purely time-like paths and purely space-like paths: What is the path yielding the shortest time-like (space-like) distances between two events and is it a space-like path or a time-like path? The same question does not make sense for purely light-like paths, as their distance is always the same, identically zero. Such shortest path are called geodesics. Again, they do not need to be unique, and their may situations arise, in which it is possible to make the path arbitrarily short or that simply no path between two events exist, or no path of a given characteristic.

Thus, this is a two-step process. The first question is, whether a purely time-like (space-like) path can be found between two events. The second is, what path creates the shortest distance. Both questions will depend, in general, on the manifold and the events in questions. Assume for the moment that a time-like path exists. The question is then which path minimizes the distance (2.12)? The structure is the same as of extremalization of a classical Lagrangian, and thus can be solved using the variational principle. Due to the square root, this is cumbersome. Doing so yields the Euler-Lagrange equations

$$\begin{aligned} 0 &= \frac{1}{d(\tau)} d_\tau \frac{g_{\sigma\nu} \frac{dX^\nu}{d\tau}}{d(\tau)} - \frac{1}{2d(\tau)^2} \frac{\partial g_{\mu\nu}}{\partial X^\sigma} \frac{dX^\mu}{d\tau} \frac{dX^\nu}{d\tau} \\ d(\tau) &= -g_{\alpha\beta} \frac{dX^\alpha}{d\tau} \frac{dX^\beta}{d\tau}. \end{aligned}$$

This equation can be recast as

$$d_\tau^2 X^\beta + \frac{g^{\beta\sigma}}{2} \left(\partial_\mu g_{\sigma\nu} + \partial_\nu g_{\sigma\mu} - \partial_\sigma g_{\mu\nu} \right) d_\tau X^\mu d_\tau X^\nu = 0.$$

This gives a conditional equation for a fixed metric. I. e., for a fixed metric, a path between two fixed events need to satisfy this condition to minimize the time-like distance. For g space-time independent, this yields the straight lines of special relativity.

This result can be brought into connection with the parallel transport (2.6). If the vector to be parallel transported is dX^μ itself, this yields

$$d^2X^\mu + \Gamma^\mu_{\nu\rho} dX^\rho dX^\mu = 0.$$

But this implies that if a vector should be transported along a geodesic then necessarily

$$\Gamma^\beta_{\mu\nu} = \frac{g^{\beta\sigma}}{2} \left(\partial_\mu g_{\sigma\nu} + \partial_\nu g_{\sigma\mu} - \partial_\sigma g_{\mu\nu}\right) = 0 \tag{2.16}$$

needs to hold. But because the transport cannot depend on the path, this implies that the Christoffel symbols need to be indeed always of this form. There is one catch, though. If (2.16) holds, then Cartan's torsion tensor (2.7) identically vanishes. The reason is the implicit assumption that the statement of transporting along the line of minimal length is the same as parallel transporting. If Cartan's torsion tensor does not vanish, then both are not the same, and moving along a geodesic is not parallel transporting a vector, and both differ by the fact that the right-hand side of (2.16) is not zero. This makes torsion a bit less obvious, and it will be picked up in section 2.4. For the moment, assume that Cartan's torsion tensor vanishes. This is the assumption of general relativity, and is experimentally consistent so far.

Intuitively, what happens is that when moving along a minimizing curve, this is not sufficient to entirely fix the orientation of a vector in directions, suitably defined, transverse to it. If the manifold has zero torsion, there is no preference. The relation (2.16) can be used to restrict the changes to be as minimal as possible. In the presence of torsion, this is not possible, as torsion will uniquely specify the necessary changes, and then (2.16) does no longer hold.

2.2 Dynamics in pure gravity

This sets the stage for general relativity. General relativity is a theory defined on a topological manifold. The topology is such that that it is locally the one of Minkowski space-time, and the induced metric has vanishing torsion and vanishing non-metricity. Thus, (2.16) holds and $D_\mu g^{\mu\nu} = 0$. The manifold is fully specified by giving $d(x, y)$ for all pairs of events.

There is a continuously infinite number of such manifolds. To make it a predictive theory requires to provide a dynamic which allows to connect a (sufficient number) of measurements of the manifold at events with predictions of the manifold elsewhere. This is provided by the Einstein equations. The Einstein equations can, in principle, be formulated entirely in terms of the distances $d(x, y)$. This will play an important role later in section 4.4. This is the so-called Regge calculus. However, in practice, this is often not convenient. It is better to formulate the dynamical principle in terms of the metric.

Given the Riemann tensor (2.9), define the Ricci tensor and curvature scalar as

$$R_{\mu\nu} = R^{\alpha}_{\mu\nu\alpha} = g^{\alpha\beta} R_{\beta\mu\nu\alpha} \tag{2.17}$$

$$R = R^{\mu}_{\mu} = g^{\mu\nu} R_{\mu\nu}, \tag{2.18}$$

respectively. Furthermore, define the Einstein tensor as

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R. \tag{2.19}$$

The dynamical equation, known as Einstein equation, is then given by

$$G + \Lambda g = 0. \tag{2.20}$$

Due to the symmetry of g, these are (in four dimensions) ten second-order partial differential equations for g. The quantity Λ is called the cosmological constant. It should be noted that at $\Lambda = 0$ the Minkowski metric $g = \eta$ is a solution to this equation. Moreover, the vanishing of non-metricity and torsion further are constraint equations, which reduce the total number of independent degrees of freedom to only two in four dimensions, and zero in less dimensions.

In principle, the only thing necessary is then to specify sufficiently many initial conditions and then solve the partial differential equations. This turns out to be even conceptually challenging. Thus, solving this equation will be postponed to chapter 3. Of course, if need be, also transfer functions will be involved.

This is it. The entirety of general relativity without matter in encoded in (2.20). It should be noted that, in the absence of matter, Newton's constant does not enter into (2.20). In particular, without cosmological constant, no dimensionful parameter appears, and general relativity is scale-invariant. The cosmological constant breaks this scale invariance and provides an intrinsic scale to the manifold.

It should be noted that the equation (2.20) is covariant constant. The second term is by requirement of the non-metricity to vanish. The first term needs tedious explicit calculation, but finally $D_\mu G^{\mu\nu}=0$ is found, basically as a consequence of the vanishing non-metricity of the metric. This actually also singles out the Einstein tensor. It is the only second-rank, torsion-free and metric tensor linear in the Riemann tensor. This singles out (2.20) as the unique covariant constant dynamical equation constant in the Riemann tensor. In a sense, it is thus the simplest dynamical equation allowing to determine a manifold itself by initial conditions. Empirically, it is indeed sufficient to describe all phenomena observed so far in general relativity, and which could be uniquely attributed to gravity.

However, it is possible to write down other covariantly constant equations, which are consistent with all observations so far. These are, however, no longer linear in the Riemann tensor, and usually also involve more than just one constant. Still, mathematically they serve the same purpose. Physically, this is an empirical choice. However, at the quantum level, they differ in general, as will be discussed in section 4.2.

2.3 Dynamics in the presence of matter

Pure general relativity (2.20) is a non-trivial theory by itself, similar to Yang-Mills theory. This will be investigated in more detail in chapter 3. However, as a theory of nature, it should also provide coupling to matter.

Covariant constancy has an important implication. For a scalar quantity, it reduces to an ordinary derivative. Thus, it implies that a scalar quantity is constant on the whole manifold. Covariant constancy ensures that this statement also makes sense for tensorial quantities. The physically same vector, i. e. suitable parallel transported, is constant everywhere, if it is covariantly constant. This is well motivated empirical, as no spontaneous generation from nothing is observed. Thus, it is necessary that matter is coupled also in a covariantly constant way to (2.20).

This requires a symmetric rank two tensor, which on Minkowski manifold is conserved in terms of the ordinary derivative. A suitable quantity is the energy-momentum tensor $T_{\mu\nu}$. This leads to the dynamical principle for general relativity with matter,

$$G+\Lambda g=8\pi\kappa T \tag{2.21}$$

where κ is an additional constant, related to Newton's constant. It should be noted that if distances are given in units of either the cosmological constant or Newton's constant to a suitable power, one of both could be eliminated from (2.21). This will not be done here, as it is useful to characterize non-trivial self-couplings[5] of gravity by Λ and of gravity to matter by κ. In particular, the limit $\kappa \to 0$ is the decoupling of matter from gravity, the situation for matter acting non-gravitational on a fixed, but potentially non-trivial, background manifold. This is the situation of, e. g., ordinary (quantum) field theory when $g = \eta$. Choosing in such a setting $g \neq \eta$ yields (quantum) field theory on curved backgrounds.

As with the left-hand side, their is no general principle to choose T. Rather, any T with the above mentioned properties will do. Eventually experiment allows to constraint and then fix T. In addition, there may exist further equations, which relate the building blocks of T with each other, e. g. Maxwell's equations. While the metric will in generally enter these equations, they are not uniquely fixed by (2.21). They again need to be supplied externally. Some examples will be given in section 2.5, where also a Lagrangian formulation for (2.21) will be given.

2.4 The vielbein formalism and torsion

At the moment, the Lorentz group is absent from the formulation. That appears particularly problematic when considering spin. It is necessary to define a notion of spin in a suitable way before being able to add non-scalar matter. Again, there are more than one possibility to do so. At the moment, and in contrast to (2.21), there is no sufficiently convincing experimental evidence to favor any solution. The one presented here is hence an example. But is one of least complexity and one which opens a straightforward way for quantization.

The key to this is the tangent space. Consider at some point in the manifold m linearly independent[6] paths $X_\mu(m(\tau_\mu))$. Build a coordinate system at this point, which consists out of the vectors defined by the derivatives at this point given by a parameter vector τ, $E_\mu = \partial_{\tau_\mu} X_\mu(m(\tau_\mu))$. The E_μ form then a basis of a vector space at this point. These are often denoted as ∂_μ, as they are really directional derivatives. The set of all such tangent spaces located at every element of the manifold is called the tangent bundle.

[5]If either would be eliminated, the same information would be obtained from the relative sizes of both fields.

[6]Note that in general it may happen that manifolds have degenerate points at which the dimensionality is lower than that of the manifold, if locally not enough independent directions exist. However, with the additional restriction of the topology allowing for a non-degenerate metric, this is forbidden.

It is possible to define a linear transformation by a matrix e_i^μ such that $e_i = e_i^\mu E_\mu$ are a set of orthonormal basis vectors in the tangent space. This requires

$$\begin{aligned} e_i^\mu e_\mu^j &= \delta_j^i \\ e_i^\mu e_\nu^i &= \delta_\nu^\mu \\ g^{\mu\nu} &= e_i^\mu \eta^{ij} e_j^\nu \\ \eta_{ij} &= e_i^\mu g_{\mu\nu} e_j^\nu \\ e_i e_j &= e_i^\mu e_j^\nu E_\mu E_\nu = e_i^\mu e_j^\nu g_{\mu\nu} = \eta_{ij}. \end{aligned} \tag{2.22}$$

that is the matrix e transforms the metric locally into the Minkowski metric η. The inverse of e_i^μ yield the basis vectors $\bar{e}^\mu = e_i^\mu dx^i$, which provide infinitesimal movements tangential to the tangent space, yielding the cotangent space. The latter is again attached at every point, and all of them together form the cotangent bundle.

The matrix e is known as the vierbein or tetrad. In dimensions different than four it is also called the vielbein. Since e is by construction invertible this allows to express any tensor either in terms of the coordinates X_μ and non-orthogonal basis vectors E_μ and the metric $g_{\mu\nu}$ or using the orthonormal basis vectors e_i and the tetrad e_μ^i or any mixtures of it.

Because of (2.15) and (2.22) it follows that under a diffeomorphism transformation

$$e_i^{\mu'} = \Lambda_\nu^\mu e_i^\nu.$$

Hence, the tetrad transforms like a vector, and especially forms a linear representation of the diffeomorphism invariance. But (2.22) allows to additionally have locally

$$e_i^{\mu'} = \lambda_i^j e_j^\mu$$

as long as

$$\lambda_i^j \eta^{ik} \lambda_k^l = \eta^{jl}$$

holds locally. But then λ is a local Lorentz transformation. Thus, the introduction of the tangent space added an additional local symmetry, a local Lorentz symmetry.

As long as no matter is present, this allows to recast (2.20) entirely in the tangent space. As a consequence, diffeomorphism symmetry is eliminated entirely in favor of local Lorentz symmetry in the tangent space. There is, at this level, no possibility to eliminate both simultaneously locally. But it is possible to shift them. Eliminating them entirely is in principle also possible, but this would turn (2.20) into an integro-differential equation.

That both can be exchanged can be traced back to the fact that the Christoffel symbols Γ are given by (2.16), and can thus ultimately be expressed in terms of the tetrads as well. This is no longer true if Cartan's torsion tensor (2.7) does not vanish. To understand this, introduce the spin connection

$$\Gamma^{\mu}_{ij} = e^{\rho}_{i} e^{\sigma}_{j} \Gamma^{\mu}_{\rho\sigma}.$$

This quantity transforms nontrivial under both the local Lorentz group and the diffeomorphism group. Expressing Cartan's torsion tensor by it yields

$$S^{\mu}_{\rho\nu} = \frac{1}{2}\Gamma^{\mu}_{ij}\left(e^{i}_{\nu}e^{j}_{\rho} - e^{i}_{\rho}e^{j}_{\nu}\right).$$

Expressing the spin connection in terms of g using (2.16) yields

$$\Gamma^{\beta}_{ij} = \frac{e^{\mu}_{i}e^{\nu}_{j}g^{\beta\sigma}}{2}\left(\partial_\mu g_{\sigma\nu} + \partial_\nu g_{\sigma\mu} - \partial_\sigma g_{\mu\nu}\right).$$

Because the torsion-free spin connection is symmetric in i and j and is contracted with an antisymmetric object in i and j in the torsion tensor, this show that indeed torsion will invalidate (2.16). But this implies that the spin connection transforms independently from the diffeomorphism symmetry, as it needs now to be non-vanishing, showing how both quantities decouple. In fact, the spin connection becomes an independent object. However, because it is introduced in the tangent space, and Einstein's equation (2.20) are completely formulated in the manifold, it is actually not a dynamical entity. It requires an alteration of these equations to make it dynamical. While this is possible classically, all possibilities come usually with an additional power of Newton's constant, making them far beyond any possibility to measure at the current time. Thus, while logically possible, space-time manifolds with non-trivial torsion are just one more possible extension of general relativity, which do not have experimental support. Thus, they will be neglected for now.

There is, however, even in a torsion-free space-time a very useful application of the spin connection, as it makes the concept of spin transparent. At the moment, only the parallel transport of tensors has been addressed using the covariant derivative (2.8). This restricts to objects which are associated with the translational symmetry. But describing particles requires spin. Spin in flat space-time is identified by representations of the Lorentz group, which has been absent so far. But this can now be addressed. Since now the Lorentz group has been recovered, the assignment of spin can be performed as in particle physics, just that the spin of matter is associated with the tangent space, rather than with the manifold.

Introducing the generators f of the Lorentz group, they can be arranged into an antisymmetric matrices of boost and rotation operators. A quantity ϕ carrying a linear representation indices r, e. g. Dirac indices, will then transform under an infinitesimal Lorentz transformation as

$$(\delta^{rs} + \omega^{ij} f_{ij}^{rs})\phi_s,$$

where ω are the space-time dependent transformation functions. This corresponds therefore to an infinitesimal Lorentz gauge transformation. Note that the contracted indices of the generators and the parameters belong to the tangent space. The entire transformation is defined there. As a consequence, there is also no way how the Lorentz transformation properties, or even the fields, can be defined in the original coordinates. The quantity ϕ_s, e. g. a Dirac spinor, makes only sense in the tangent space. To eliminate again the tangent space and return to the manifold requires to build Lorentz-invariant quantities, e. g. $\phi_s^\dagger \phi^s$, which again are well-defined outside the tangent space. It should be noted that the Lorentz indices r and s are not tangent space indices, and can therefore not be translated using the vierbein.

A further consequence is that the covariant derivative now needs to take care of the comparability of the local Lorentz group. Thus, the covariant derivative becomes

$$D^\mu_{rs}\phi^s = \left(\delta^{rs}\partial^\mu + \Gamma^\mu_{ij} f^{ij}_{rs}\right)\phi^s. \tag{2.23}$$

The so-created object is now a tensor, with one index from the coordinate space and one from the tangent space. It should be noted that, despite that ϕ^s does not carry a coordinate index μ, it does change under a diffeomorphism transformation, as it depends on coordinates. In particular in infinitesimal form

$$\Lambda\phi = (1 + \epsilon^\mu(X)\partial_\mu)\phi(X)$$

where ϵ is the local translation of (2.13), generated by the derivative as the generator of local translation. The affine connection does not appears because ϕ is scalar with respect to the coordinate system on the manifold. Of course, mixed tensors would be possible, but do not add further conceptual complications.

The usual spin of flat-space (quantum) field theory arises, because the metric g becomes constant, $g = \eta$. As a consequence, so do the tetrads, $e^a_\mu = \delta^a_\mu$. Thus, both the Lorentz transformation and the translations of (2.14) become global transformations.

However, a global Lorentz boost in corodinates, Λ_μ, will alter the tetrads, $\Lambda^{\mu\nu}e_\mu$. To mainatin $e_\mu^a = \delta_\mu^a$, this requires also a compensating global Lorentz transformation in the tangent space, $\left(\Lambda^{-1}\right)^{ab} e_b$. Therefore, objects carrying a Lorentz index in local othorgonal tangent space need to transform in the same way as those carrying a tensor index in the coordinate space. After all, the teatrads arose from the introduction of an local coordinate change, and this needs to be maintained once having only global rotations alone. Thus, both groups are thereby reduced to a diagonal subgroup, thereby yielding the single Poincaré group of special relativity.

2.5 Lagrangian formulation

For the purpose of a path integral formulation, as well as for manifest coordinate-independent approaches to classical general relativity, it is useful to recast the tensorial Einstein's equation (2.20) in a Lagrangian form. Einstein's equation will then reemerge from a variational principle. It is thus very similar to the action of Maxwell theory, though with the added complication of diffeomorphism symmetry.

The probably most important change is that d^dX is not an invariant volume element. Because diffeomorphism symmetry allows also for a rescaling of coordinates, volume can get rescaled as well. This change needs to be offset. In fact, this can be compensated by the covariant volume element $d^dX\sqrt{-\det g}$. This quantity transforms like an inverse density. Thus, using a Lagrangian density $\mathcal{L}$, the action takes the form

$$S = \int d^dX\sqrt{-\det g}\mathcal{L}. \tag{2.24}$$

A suitable Lagrangian density is given by

$$\mathcal{L} = \frac{1}{2\kappa}\left(R + l\right)$$

where Newton's constant is introduced as a rescaling for latter convenience. Thus, the total action is the manifold-integral of the sum of curvature and the cosmological constant. Note that in Minkowski space-time $R = 0$ and $l = 0$, and the action vanishes identically. Thus, it is necessary to first perform the variational principle before making a choice to obtain Einstein's equation, as usual. This Lagrangian is known as the Einstein-Hilbert Lagrangian. Furthermore, it should be noted that the Lagrangian is quartic in the fields and contains derivatives up to order two. It is thus the equivalent of the field strength tensor squared of Yang-Mills theory, and demonstrates that even pure gravity is self-interacting, and thus a non-trivial theory.

Since R is a curvature density, this implies that Einstein's equation extremalizes total curvature on the manifold. At $l = 0$, a possible extrema is zero, and thus Minkowski space-time is a solution to this problem. Solutions at non-zero l will be discussed in chapter 3.

Adding matter yields

$$\mathcal{L} = \frac{1}{2\kappa}(R + l) + \mathcal{L}_M \tag{2.25}$$

where $\mathcal{L}_M$ is the minimally coupled matter Lagrangian. Similar to the flat space-time gauge theories, this implies for a free, massless scalar field

$$\frac{1}{2}\partial_\mu\phi\partial^\mu \to \frac{1}{2}g^{\mu\nu}D_\mu\phi D_\nu\phi = \mathcal{L}_M, \tag{2.26}$$

where the covariant derivative is given by (2.8). Of course, for a scalar field this immediately reduces back to the ordinary derivative. In general, a coordinate tensor field will involved the affine connection. Moreover, if the field has spin, e. g. fermions, this requires to further invoke a term (2.23), as required by the phenomenology.

This implies that even the simplest matter field automatically couples to gravity. This happens due to two effects. One originates from the covariant volume element in (2.24). The other appears from the contraction of the derivatives. Thus, it is impossible for the metric not to interact with matter, even if the matter would be massless as in the present case. Or, matter always interacts with the metric. It is this universal interaction which is known as gravity, as will become evident in section 3.2. Gravity is really just an ordinary gauge interaction, but which couples even to derivatives.

Note that because of dimension also, e. g., a term proportional to $R\phi^2$ can be added to the action. This would imply a direct coupling of the scalar to curvature, and would constitute a mass-like term. Curvature acts like a (space-time-dependent) mass, on top of any tree-level mass. However, any mass term already couples also to the metric by the covariant volume element, and thus gravitates as well. This is the origin of the idea that gravity acts upon mass, while it really couples even to self-interaction terms like ϕ^4, if they are present. Note that this coupling occurs by gauging the argument of the fields, rather than the field itself, which makes it look so different than from ordinary interactions. But since coordinate-dependency can be considered to be a representation of translation symmetry, it is actually not that different, if one thinks of a gauge symmetry with a continous index in an infinite-dimensional representation space rather than the usual finite-dimensional ones.

Having a Lagrangian like (2.25) yields two coupled equations of motion. In their derivation, it is necessary to observe that the action is actually varied, not the Lagrangian alone. As a consequence, the factor $\sqrt{-\det g}$ needs to be included in the Lagrangian equations of motion. One equations following is then Einstein's equation with matter (2.21), where T now becomes

$$T_{\mu\nu} = \frac{\delta\sqrt{-\det g}\mathcal{L}_M}{\delta g^{\mu\nu}} \stackrel{(2.26)}{=} -\frac{2}{\sqrt{-\det g}}\frac{1}{2}\partial_\mu\phi\partial_\nu\phi,$$

where it was used that

$$\delta g = -g g_{\mu\nu}\delta g^{\mu\nu}$$

holds. Thus, the energy momentum tensor stems from the variation of the matter Lagrangian with respect to the metric. In the Minkowski case, this would be the energy-momentum tensor. Of course, if the matter Lagrangian is not that of a scalar field, the affine connection appears, which depends in general relativity on the metric, and its derivatives. Then

$$T_{\mu\nu} = -\partial_\rho\frac{\delta\sqrt{-\det g}\mathcal{L}_M}{\delta\partial_\rho g^{\mu\nu}} + \frac{\delta\sqrt{-\det g}\mathcal{L}_M}{\delta g^{\mu\nu}}.$$

Likewise, the equation of motion for the matter field becomes

$$0 = \partial_\mu\frac{\delta\sqrt{-\det g}\mathcal{L}_M}{\delta\partial_\mu\phi} - \frac{\delta\sqrt{-\det g}\mathcal{L}_M}{\delta\phi} \stackrel{(2.26)}{=} \partial_\mu\left(g^{\mu\nu}\sqrt{-\det g}\partial_\nu\phi\right). \tag{2.27}$$

At $g = \eta$, this yields the usual plane-wave equation $\partial^2\phi = 0$. However, if g varies, this implies a term sourcing changes in the field as

$$\partial^2\phi = -\left(\partial_\mu g^{\mu\nu}\sqrt{-\det g}\right)\partial_\nu\phi$$

which depends on the changes of the field itself. This equation also holds true if the matter does not backcouple, i. e. Einstein's equation is dropped. Hence, in a static curved space-time even a massless particle will not follow straight lines, but rather move along a line determined by this non-trivial term. Conversely, the presence of a matter field sources a change in the metric, and thus alters the manifold.

Note that both (2.25) and (2.27) are field equations. The metric determines the topology. The pure-gravity equation (2.20) is a self-consistency equation for the topology of the manifold. The manifold itself is given by its set structure. This determines which events have which neighboring relations. In particular, this determines the overall global structure, e. g. of a sphere or a torus[7]. The metric determines the topology, i. e. what is the distance between neighboring events, and thus arbitrary (geodesic) distances on the manifold. The overall manifold structure thus determines which events are neighboring to each other, and given this input, Einstein's equation (2.20) determines the distances. Einstein's equation cannot change which events are neighboring in terms of the sets. However, it can create a topology where the values of the distances do not reflect neighboring relations, i. e. events close by in terms of overlapping sets can have, relative to other pairs of events, very large or very small distances. This is even true for Riemannian manifolds, where the distances are positive semi-definite. The characteristic feature of the neighboring relation is only that there exists always, at least, an infinitesimal neighborhood, in which the distances reflect the neighboring relation, in terms of a Minkowski metric[8].

Once matter is added, at every event a value for the matter field is obtained. The presence of the matter alters the topology, as distances change according to (2.25). However, at the same time the field is affected by (2.27). Thus, the two coupled equations will eventually settle for a self-consistent set of distances and field amplitudes. As both are partial differential equation, this solution will depend on both an initial condition, to be discussed in more detail in section 3.1, as well as the manifold structure in terms of neighboring relations.

Note that the concept of a point particle has not been discussed. The problem is that the distributional character of a point particle's energy momentum tensor makes very challenging to understand the back reaction of the particle on itself. E. g., a massive particle will move, and thereby modify the metric. This changes may spread at the speed of light. Thus, the point particle will eventually interact with this change. That is very similar to the case of the back reaction of an electrically charged point particle in electromagnetism. But the technical challenges involved amplifies in the general relativistic case. Thus, point particles will in the following only be considered as probe particles, i. e. without any (relevant) change on the metric. When using fields,

[7]Sometimes this is also called the topology of the manifold. To avoid confusion with the topology in terms of distances this terminology is not used here. Furtheremore, many neighboring relations can be deformed such that only the number of holes, the so-called genus, of an isomorphic surface is relevant. This is e. g. the case for sphere and torus. A rotational ellipsoid will work the same as the sphere. However, examples like the Klein bottle in three dimensions, which is also a valid manifold in terms of neighboring relations, shows that this is not generally simple.

[8]It may appear at first sight that then the distances should globally be such. However, due to the possibility for transfer functions, this can change quickly radically for finite distances.

this problem does not arise, as the field are not distributions. These are just ordinary coupled partial differential equations.

2.6 Energy, Fourier-space and physical observables

In the context of wave equations switching to Fourier space is often advantegous. As a consequence, usage of momenta is quite common in physics. This is also true already in special relativity, where four momentum conservation becomes a central tenet. This compensates for the loss of individual meaning of energy and three-momentum.

The situation becomes more involved in general relativity. Switching to momentum space is done by a Fourier transformation. However, the exponential factor $\exp(i\eta^{\mu\nu}x_\mu p_\nu)$ needs now to be replaced by $\exp(ig^{\mu\nu}X_\mu P_\nu)$ to ensure that the argument of the exponential factor transforms like a scalar under coordinate transformation. In addition, the volume has to be adapted, leading to

$$f(P) = \int d^dx\sqrt{-\det g} f(X)e^{ig^{\mu\nu}X_\mu P_\nu}.$$

While this is formally possible, though in general will be technically involved due to the space-time dependence of the metric, this has far-reaching conceptual problems. Because exactly the requirement of being scalar requires that under a coordinate transformation $P \to \Lambda P$ becomes an event-dependent transformation. Thus, the value of the four momentum as whole, not only of its components, because malleable by diffeomorphism transformations. Thus, just like the vector-potential in electromagnetism or the electric field in Yang-Mills theory, it can no longer be a physical observable. Only the four-momentum squared $P_\mu g^{\mu\nu}P_\nu$ remains invariant, just as distances are.

This is not entirely surprising, after all also coordinates lose their meaning as something, which is at least in principle can be made real. In a sense, in a manifestly covariant description, this does not change too much. After all, already in special relativistic theories quantities should be only dependent on squared quantities. However, The problems now appears when considering a quantity like the scalar product between two different four-momenta. The space-time dependent transformation Λ will not transfer into a momentum-dependent transformation in Fourier space. Likewise the metric needs to be transformed.

What this means can already be seen in position space. Consider the invariant distance, but expand it as,

$$ds^2 = (X_\mu - Y_\mu)g^{\mu\nu}(X_\nu - Y_\nu) = X^2 + Y^2 - 2X_\mu g^{\mu\nu} Y_\nu, \tag{2.28}$$

where it was used that the covariant derivative of the metric vanishes. The full expression is necessarily invariant under a coordinate transformation, by construction. However, because the coordinate transformation is local, the difference $X-Y$ is no longer invariant itself, and the remainder is compensated by the change in metric. As a consequence, an expression like X^2 is not itself invariant, as here the metric changes. Only the combination of all three terms in (2.28) yields the invariance. This especially implies that the term XY is not invariant. Hence, scalar products, and thus angles, are not necessarily an invariant under diffeomorphism transformation. Therefore, a dependency on special Poincaré-invariant quantities is not, in itself, sufficient to be diffeomorphism invariant.

As a consequence, only quantities depending on distances can be invariant. And this makes geodesic distances even more special. They at least have a chance to define path-independent observables, if the geodesic distances are unique.

This does not mean that Fourier space is useless. However, it needs to be considered an auxiliary when calculating eventually diffeomorphism invariant quantities, very much like the vector potential in classical electromagnetism. Of course, once approximations are introduced, it becomes as challenging as in other theories to maintain this invariance.

These issues can, to some extent, be ameliorated in practical cases. This will be discussed in section 3.9.

Chapter 3

Special solutions

Special solutions, like plane electromagnetic waves in classical electrodynamics, are both useful, but also challenging. The main obstacle is again that there is no arena for the special solutions to exist in. Electromagnetic plane waves can be studied, because they are existing in an enclosure of space-time. Special solutions in general relativity do not have such an enclosure. Therefore, in principle, always the whole space-time needs to be considered.

As will be seen, in the absence of matter, solutions can become maximally symmetric, see section 3.3. Some special solutions will address in some sense localized configurations, like black holes in section 3.5. They allow to consider these configurations to be embedded in a specific sense in the maximally symmetric cases without altering their overall character too much. This is probably the closest analogue to the electrodynamics case.

However, other special solutions cannot be considered such. Most prominent of them are those which describe cosmology in section 3.4. Here, everything is, in a sense, special. This makes it also hard to wrap one's mind around their specialness.

In general, most special solutions which do not involve small curvature yield substantial deviations even from the logic of special relativity. This makes them very non-intuitive and difficult to grasp.

3.1 The initial value problem

Einstein's equation, without matter (2.20) or with matter (2.21), are partial differential equations. They are supplemented by conditions of vanishing torsion and covariant constancy. They are thus subject to constraint equations. These reduce the independent degrees of freedom from 16 (in four dimensions) down to 6. Still, in the end only two are independent. This leads to a freedom in the choice of the components, similar to the gauge freedom of classical electrodynamics. Imposing, e. g., the Haywood gauge[1] $\partial^\mu g_{\mu\nu}$ introduces four more constraints, reducing it to the necessary two independent degrees of freedom.

As if solving partial differential equations with constraints is not difficult enough a task[2], there is another severe problem. These differential equations involve expressions like $g^{\mu\nu}\partial_\mu\partial_\nu g_{\rho\sigma}$.

In flat space-time $g = \eta$ this Laplace-Beltrami operator $g^{\mu\nu}\partial_{\mu\nu}\partial_\nu$ has on non-constant modes a positive definite spectrum. In a general metric, this is no longer necessarily true. Likewise the various possible other second-order differential operator appearing are no longer of this type. Since many uniqueness and existence proofs do, however, rely on these properties, they are not applicable to the Einstein's equation. As a consequence, no complete theory of the solution manifold of the most general case of general relativity with (ordinary) matter is available. As a consequence, it is not even clear, which kind of initial values are needed to find a unique solution.

This problem can be largely remedied by restricting the possible initial data. However, the condition will only be possible to realize on certain manifolds, reducing the possible manifolds.

The restriction is that the initial values for the metric g describe a space-like submanifold of one dimension less. I. e., for a submanifold, which is given in terms of a subspace of the underlying $\mathbb{R}^d$, all invariant distances are of the same sign, space-like, and for different elements non-zero. In a visual picture, it is possible that a space-like slice through the universe is possible and specified. If this is the case, it can be shown that the vacuum Einstein's equation evolve such that the whole manifold is separated into such space-like submanifolds, i. e., every element of the manifold belongs to precisely one such space-like sub-manifold. Moreover, any distances between two sub-manifolds are never space-like, and for every element on a given sub-manifold a time-like separated element on a neighboring, in the manifold-sense, sub-manifold can be found. In a visual

[1] Viable gauge conditions, i. e. those for which for every diffeomorphism orbit at least one solution is guaranteed to exist, are not necessarily simple to construct.

[2] An approach like in classical mechanics to incorporate the constraints by switching to generalized coordinates is in general challenging due to the transfer functions. This will usually change the character of the equations to integro-differential ones.

picture, the universe is sliced into space-like slices, separated by time-like distances. This is called a foliation. It is closest to the conventional idea of a universe. While there is no local unique time, there is a global universe time, which essentially counts the slice of the stack. Such initial conditions are therefore called physical, and the decomposition into space-like submanifolds a foliation of the manifold. In a sense, this is also the simplest decomposition into a Riemannian part and a non-Riemannian part of the manifold.

Of course, the space-like submanifolds are still Riemannian manifolds, and can be very involved, and especially still require a non-trivial atlas. An example would be the surface of a four-dimensional sphere, which satisfies these criteria. It is also possible that the submanifold is not simple, and thus describe separate universes. Still, this allows to guarantee a unique solution. If matter is added, this requires it to be ordinary matter, e. g. having a positive energy density in terms of the energy-momentum tensor. Otherwise, such a source term would spoil the evolution for the same reason as arbitrary initial conditions would do.

As a submanifold, such space-like hypersurfaces can be in turned described as before, just with one dimension less. Thus, there exist also a metric, which is called the induced metric h. Of course, the induced metric is fully known in terms of the metric g of the whole manifold. This observation allows a convenient description of the foliation. The induced metric requires to obtain the full metric the difference between neighboring points, in the sense of the full manifold, of two different space-like submanifolds. The connecting quantity is called the lapse function or, when expressed correspondingly, the lapse vector. It encodes locally the time-like separation of each space-like slice, and thus the local passing of time in a suitable rest frame.

In practice, this implies for a curve $X(\tau)$ its tangent vector can be decomposed as

$$\partial_\tau X(\tau) = \alpha n + \beta$$

where n is a future time-like vector of unit length in the tangent space. The function α is then called the lapse function and β, being a vector in the tangent space, is called the shift vector. It is therefore tangential to the space-like hypersurface. This therefore decomposes a direction into one which connects neighboring space-like hypersurfaces, $n\alpha$, as well as a component which describes the shift inside the hypersurface when changing from one to another. Especially, this implies that $n\alpha$ describes the direction in which space-like coordinates can be found, which do not change. Physically, moving along $n\alpha$ is thus the closest resemblance to a worldline in a rest frame.

This allows to locally decompose the invariant length between two point on neighboring hypersurfaces connected by this path as

$$\begin{aligned} ds^2 &= -(\alpha^2 - h^{rs}\beta_r\beta_s)dx_0^2 + 2h^{rs}\beta_r dx_s dx_0 + h^{rs}dx_r dx_s \\ &= -\alpha^2 dx_0^2 + h^{rs}(dx_r + \beta_s dx_0)(dx_s + \beta_s dx_0), \end{aligned}$$

where the $(d-1)\times(d-1)$ induced metric h in the spatial hypersurface appears as well as r and s label the components with the space-like hypersurface. In particular, this shows that along a pure time-like distance only the lapse function contributes, while $dx_r + \beta_r dx_0$ essentially provide the orthogonal, in the sense of this metric, displacement.

Switching to the orthonormal coordinates of a vielbein, this yields that $e^0 = n$, i. e. one of the vielbeins is just the time-like future direction between the two hypersurfaces. The spatial ones decompose then the induced metric as $e^i_r\delta^{rs}e^j_s = h^{ij}$, as the hypersurface is a Riemannian submanifold.

This particular set of coordinates is well adapted to the foliation of space time. It is therefore well suited to solve the initial value problem. However, choosing suitable coordinates in the remainder space-like hypersurface may still be a non-trivial issue. Also, transfer functions may still be needed.

Another concept, which is useful in the context of discussing the initial value problem, is the concept of Killing vectors and fields. Diffeomorphism symmetry and local Lorentz symmetry are, in a sense, non-existing. They only arise as a consequence of the introduction of coordinates and the tangent space. They are not necessary to describe the manifold. However, the manifold can have additional symmetries. Since the manifold is entirely specified in terms of neighboring relations and distances, any such symmetries can only exist in terms of these, and thus in terms of the distances. Thus, they are also called isometries in this context.

An isometry is obtained if there are path in the manifold, which can be displaced in the manifold, without altering the distances between neighboring points along the curve. If there is a displacement prescription such that this can be achieved for the whole manifold, this is an isometry. It basically expresses that a manifold looks the same when one 'rotates' it along a certain direction. Isometries therefore describe global properties of the manifold. E. g. a torus or a sphere do have this property by rotating around their symmetry axis. This operation does not change distances along curves. Thus, these two manifolds have isometries. There can be multiple such isometries.

While isometries can be defined entirely in terms of the topology of the manifold, it is usually more convenient to utilize coordinates. This is then achieved in terms of a Killing field. An isometry is obtained if, in a fixed coordinate system, the metric will not change under a displacement. As a consequence, for an isometry, there exists always a coordinate system such that the metric becomes independent of a suitably chosen coordinate. In general, this will be a directional derivative

$$\partial_\tau X^\mu \partial_\mu g_{\alpha\rho} = 0. \tag{3.1}$$

Of course, isometries of submanifolds are also possible. In the case of a foliation particularly interesting are isometries determined using the induced metric h. Thus, the direction of an isometry is given in terms of the base vector obtained from $\partial_\tau X_\mu = \xi_\mu$. The directional derivative $\xi_\mu \partial^\mu$ therefore creates a direction under which distances are not changing. This can be immediately seen as the change is independent of the path parametrization, and thus

$$ds^{2'} = (g^{\mu\nu} + \xi_\rho \partial_\rho g^{\mu\nu}) \frac{d(X_\mu + \xi_\mu)}{d\tau} \frac{d(X_\nu + \xi_\nu)}{d\tau} = ds^2$$

by virtue of (3.1).

If ξ indeed determines an isometry and is thus a Killing vector can be tested using the Killing equation,

$$D_\nu \xi_\mu + D_\mu \xi_\nu = 0. \tag{3.2}$$

This is a covariant statement. It is therefore admissible to choose a coordinate system, if the calculation is performed exactly, and prove it in this system. Choosing a coordinate system in which the Killing vector is constant, i. e. one of the coordinate axis is taken to be identically to the direction of the Killing vector, yields

$$D_\mu \xi_\nu = g_{\mu\alpha} \Gamma^\alpha_{\nu\sigma} \xi^\sigma = \frac{1}{2} \left(\xi^\sigma \frac{\partial g_{\mu\sigma}}{\partial X^\nu} + \xi^\sigma \frac{\partial g_{\mu\nu}}{\partial X^\sigma} - \xi^\sigma \frac{\partial g_{\nu\sigma}}{\partial X^\mu} \right) = \frac{\xi^\sigma}{2} \left(\frac{\partial g_{\mu\sigma}}{\partial X^\nu} - \frac{\partial g_{\nu\sigma}}{\partial X^\mu} \right).$$

This is hence an antisymmetric quantity, thus implying Killing's equation.

Killing vectors therefore provide a particularly suited to form a basis. Moreover, they imply the existence of a conserved quantity. Given some curve $X(\lambda)$, then

$$p_\xi = \xi^\mu d_\lambda X_\mu$$

is constant along the curve. This can be seen by taking a derivative with respect to λ,

$$d_\lambda p_\xi = d_\lambda \xi^\mu d_\lambda X_\mu + \xi^\mu d_\lambda^2 X_\mu = 0$$

Choosing a coordinate system with ξ_μ a basis vector and where the curve is linearly dependent on λ, this expression vanishes. Thus, the component p_ξ is indeed conserved.

Choosing the same system of coordinates to introduce a Fourier transformation allows to identify this quantity as the momentum component along the direction of the Killing vector, which is conserved.

Physically, this is not entirely surprising. After all, an isometry is really a generalized global translation symmetry, which entails a conserved generalized momentum. More surprising is probably that this is not a build-in symmetry, like in classically mechanics. If the solution to the dynamical equations yield a manifold with Killing vectors, the system has dynamically this symmetry.

It should be noted that the maximal number of Killing vectors is limited. Killing's equation (3.2) is a differential equation for the Killing vectors. Since $D^2\xi = 0$ because of the covariant constancy of the metric, this implies that the Killing vectors can be obtained as solution to a second-order partial differential equation, which has to obey antisymmetry of the first-order derivatives. For an n-component field, this yields at most n components with at most $(n^2 - n)/2 = n(n-1)/2$ independent derivatives. Thus, there can be at most $n(n+1)/2$ independent Killing vectors.

3.2 Identifying gravity

It has been mentioned several times that general relativity is basically the gravity from Newtonian physics. As noted before, Newtonian physics, or even special relativity, need to have flat space-time. It is therefore a requirement on the manifold structure. However, in many cases it is not necessary to consider the case that the whole of space-time has this feature, but often it is sufficient to consider only some small patch to be almost flat. Of course, by construction, there always exists an infinitesimal neighborhood where this is possible, but here a somewhat larger patch is needed.

Assume that in some patch, e. g. earth, the metric in a suitable coordinate system can be written as

$$g = \eta + \gamma$$

where for units in which the metric is dimensionless $|\gamma_{\mu\nu}| \ll 1$ holds. Of course, h itself is not a metric, while η is, as is $\eta + h$. Especially, lowering or raising of indices will not work with γ.

Neglecting quadratic and higher power terms of γ simplifies the Riemann tensor (2.9) as thus the Ricci tensor (2.17) which is needed to determine the Einstein tensor (2.19) and thus the equation of motions for matter. The Ricci tensor takes the form

$$R_{\mu\nu} = \partial_\lambda \Gamma^\lambda_{\mu\nu} - \partial_\nu R^\lambda_{\lambda\mu} + \mathcal{O}(\gamma^2). \tag{3.3}$$

Considering further that the aim is to establish the relational to Newton's law of gravity, it is acceptable to neglect time variations of γ.

Newton's law of gravity can be formulated as

$$\partial_i^2 \phi = 4\pi G_N \rho,$$

where ϕ is the gravitational potential, and ρ is the matter density. $G_N = \kappa/(8\pi)$ is then Newton's constant in its usual form. To make contact requires therefore to consider the matter density. The matter density is T_{00} for a classical continuum mechanics system. Thus

$$\partial_i^2 \phi = 4\pi G_N T_{00}. \tag{3.4}$$

Setting $\gamma_{00} = 2\phi$ and combining (3.3) and (3.4) with (2.21) yields

$$\partial_i^2 \frac{\gamma_{00}}{2} = \partial_i^2 \phi = 4\pi G_N \rho - \Lambda. \tag{3.5}$$

Thus, indeed the metric yields the classical gravitational potential. Especially, if the backreaction of the matter on the gravitational potential is neglected, which is justified with γ being small, this is exactly Newton's law of gravity, up to the appearance of the cosmological constant. The latter acts like a constant matter (or energy) density. Only measurement can decide its value. For a large value, it would change Newton's law of gravity. However, its value is measured to be so small, that it can safely be neglected in this context. This also justifies the identification of κ with Newton's constant.

This also shows that the metric is essentially the gravitational potential. There are, however, two important observations to be concluded from this.

First, even at weak gravitational field the equation (3.5) should not be read as the equation of motion for matter, i. e. the equivalent of Newton's second law. This is actually given by (2.27) as Lagrange's second equation. Rather, equation (3.5) and (2.27) are needed to be solved together. In fact, (3.5) can be recast, neglecting Λ, as

$$\phi(\vec{r}') \sim \int d^3r \frac{\rho}{|r - r'|}, \tag{3.6}$$

showing that the gravitational potential, or the metric, is determined by the mass density. In turn, entering γ into (2.27) then shows how the matter density changes under the influence of gravity. The fact that (3.5) is usually given as Newton's gravitational law is that a probe particle, e. g. earth, is considered in the gravitational field of a much larger body, e. g. the sun, who by virtue of (3.6) determines essentially the gravitational potential, and then (3.5) is only considered close to the location of the probe particle. Of course, self-consistently, this will coincide with the solution of (2.27) in this case.

Second, in the whole process never a mention of an inert mass and a gravitational mass was made. This is not due to a tacit identification of both. Rather (2.21) and (2.27) together originate as dynamical equations from a Lagrangian which does not necessitates any such distinction, in fact does not even make it possible. Inertial mass and gravitational mass can only be distinguished if equations (2.21) and (2.27) are not solved simultaneously, and then the masses in either of them can be identified such. Thus, the fact that they form a coupled equation derives from a single Lagrangian establishes that there is not independent existence of either.

3.3 Maximally symmetric solutions

Given the usual logic of theoretical physics, the first example of solutions to the initial value problem will be the pure general relativity case (2.20) with the aim to find maximally symmetric solutions. Thus, the aim is to maximize the number of Killing vectors. This could be at most four, or in general the number of dimensions, corresponding to the fact that each Killing vector describes a coordinate of which the metric becomes independent.

As Einstein's equation (2.20) stands, it allows for a rescaling of the cosmological constant. In the absence of any other units, it is always possible to measure the coordinates in suitable powers of it. Thus, Einstein's equations can always be rescaled such that the cosmological constant is either zero or ± 1. Being larger or smaller than one then implies that the distances are small or long compared to the appropriate power of the cosmological constant, which gives the natural scale. But in absence of another scale, this implies that the three cases cannot continuously be deformed into each other. There is always a scale above or below the characteristic one. Also, since Einstein's equations are second-order equation there is no possibility to scale out the sign of the cosmological constant. Thus, it is necessary to treat the three cases independently.

The simplest case arises when the cosmological constant vanishes. There is no characteristic scale in the system. Moreover, employing maximum symmetric requires the metric to be constant, and scalelessness implies that all components have equal size. This yields $g = \eta$ as solution, as the Einstein tensor (2.19) vanishes at $g = \eta$, since the Riemann tensor vanishes as it only depend on derivatives of the metric. Hence, there is a Killing vector for every direction, yielding back momentum conservation of special relativity. In addition, the six Lorentz generators complete the list of the possible ten independent Killing vectors in four dimensions.

At non-vanishing cosmological constant the situation is more involved. Still, the aim is to find a solution with the maximal number of Killing vectors. It is best to do so in a suitable fixed basis. The aim is still to find a foliated space-time. Moreover, to achieve maximal symmetry the spatial part needs to be likewise highly symmetric. In a suitable coordinate system, the metric then takes the Friedmann-Lemaitre-Robertson-Walker (FLRW) form

$$g_{\mu\nu} = -dt^2 + a(t)d^3\Sigma. \tag{3.7}$$

where $d^3\Sigma$ describes a homogeneous and isotropic spatial hypersurface. The only possibilities in general relativity for such a structure is

$$d^3\Sigma = \frac{dr^2}{1-kr^2} + s_k(r)^2 d^2\Omega = \frac{dr^2}{1-kr^2} + s_k(r)^2(d\theta^2 + \sin^2\theta d^2\phi),$$

where θ and ϕ are the usual polar and azimuthal angle. The quantity k determines the curvature of the hypersurface, yielding

$$\begin{aligned} s_{k>0} &= \frac{1}{\sqrt{k}} \sin\left(r\sqrt{k}\right) \\ s_{k=0} &= r \\ s_{k<0} &= \frac{1}{\sqrt{k}} \sinh\left(r\sqrt{-k}\right). \end{aligned}$$

Thus, positive values determine a spherical hypersurface, zero is a flat hypersurface, and a negative value is hyperbolically shaped. It is often customary to rescale $r \to rk^{-\frac{1}{2}}$, and thus distances are measured in units of $k^{-\frac{1}{2}}$. Then k becomes discrete with the three options ± 1 and 0. Note that throughout it will be assumed that the spatial hypersurface has trivial topology, i. e. no complicated boundary conditions, is simply connected and has no holes. The choice of k is part of the initial conditions, as is the fact that the spatial part has no non-trivial boundary structure.

Inserting this form into the definitions yield the curvature scalar (2.18)

$$R = 6\left(\frac{\partial_t^2 a}{a} + \frac{\partial_t a}{a^2} + \frac{k}{a^2}\right)$$

and thus the curvature is both time-dependent and depends on k. In fact, in a static universe, $a(t) = a_0$, k determines entirely the curvature. If the spatial hypersurface is a point, $a = 0$, the curvature diverges.´

The factor a is determined by the solutions of Einstein's equation (2.20), and thus by the cosmological constant,

$$\frac{\partial_t^2 a}{a} = \frac{\Lambda}{3} \tag{3.8}$$

which is solved by

$$a(t) \quad = \quad a_+ e^{-\sqrt{\frac{\Lambda}{3}}t} + a_- e^{\sqrt{\frac{\Lambda}{3}}t}$$

where the prefactors are again determined by the initial conditions. This implies that $a(t)$, which plays the role of a scale factor of the metric, is exponential for $\Lambda > 0$, but periodic for $\Lambda < 0$. I. e. in the first case, called de-Sitter, the universe will have distances changing exponentially with time t. In the second case, called anti-de Sitter, the distances will change periodically. Note that in both cases the spatial structure can be the same, and the prefactor only regulates the time-dependence of distances.

To quantify the behavior of the distances the Hubble parameter is introduced as

$$H(t) = \frac{\partial_t a}{a}, \tag{3.9}$$

it measures the relative change of distances as a function of t. Since both cases describe foliated space-time, the time coordinate is, in principle, eternal. Both solutions are indeed maximal symmetric. This can be seen by the fact that both de Sitter space-time and anti-de Sitter space-time can be described in terms of an embedding in a higher-dimensional Minkowski space time as the surface of a hyperboloid, which differ only by whether time is a closed line around the hyperboloid (anti-de Sitter) or along the open direction of the hyperboloid (de Sitter), with space playing the opposite role. On such a surface there are, of course, the one dimension less symmetry group, and thus giving again a ten-dimensional group, giving ten Killing vectors. Although their interpretation is now less forward than flat Minkowski space-time.

Note that also flat Minkowski space-time fits into this structure, yielding in fact also the two solutions $a(t) = a_0 t + a_1$, rather than just $a(t) = 1$. However, since this is only a time-dependent rescaling, this does not alter the result. Note that the curvature is in general no longer time-independent, but will depend on the interplay and relative size of the different constants.

3.4 Big-bang solutions

The solutions in 3.3 give pure gravity solutions. The natural question is what happens once matter is introduced into the system by the set of coupled equations (2.21) and (2.27). The relevant quantity is then the energy-momentum tensor in (2.21). Assuming for the moment that matter interactions as described by (2.27) are such that they quickly establish local equilibrium, the coupling between both equations will be small, and the energy momentum tensor will have the same isotropy and homogeneity as space-time itself. In a suitable coordinate system the energy-momentum tensor will be of the form $T_{\mu\nu} = \mathrm{diag}(\rho, p, p, p)$, where $\rho(t)$ is the matter density and $p(t)$ is the pressure. Both will be spatially constant on the homogeneous space-like hypersurfaces. The vast difference in coupling strengths between Newton's constant and the typical couplings of the other interactions justify this approximation. In a universe with different values, this would probably not be justified.

The matter density and the pressure are not independent, but related by the equation of state of the matter derived from (2.27). This yields the two coupled equations

$$\partial_t \rho = -3\frac{\partial_t a}{a}(\rho + p^2) \tag{3.10}$$

$$\frac{\partial_t^2 a}{a} = -\frac{4\pi G_N}{3}(\rho + 3p) + \frac{\Lambda}{3} \tag{3.11}$$

where (3.11) is derived from (2.21), and is the version with matter of equation (3.8).

Note that replacing

$$\begin{aligned} \rho &\to \rho + \frac{\Lambda}{8\pi G_N} \\ p &\to p - \frac{\Lambda}{8\pi G} \end{aligned}$$

in (3.10-3.11) would eliminate the cosmological constant. This is equivalent to say that in this setting the cosmological constant behaves like matter satisfying the equation of state

$$\rho_\Lambda = -p_\Lambda \tag{3.12}$$

i. e. like a positive matter density which exerts a negative pressure for a positive cosmological constant, and negative matter density which exerts a positive pressure for a negative cosmological constant. Thus, the cosmological constant tends to blow up the universe in case of the observed positive value.

It follows that the Hubble parameter (3.9) is a convenient quantity for the following it. It takes the form

$$H^2 = \frac{8\pi G}{3}\rho - \frac{k}{a}.$$

This can be used to define a critical density

$$\rho_c = \frac{3H^2}{8\pi G}$$

which implies that the quantity

$$\Omega = \frac{\rho}{\rho_c}$$

is, for k/a^2 negligible, a measure for the fate of the universe. If it is larger than 1, the normalized change of rate yields that the universe will eventually collapse. If it is equal one, it will reach an asymptotic size. And if it is smaller than one, the universe will expand forever. Current measurements for our universe strongly suggest $k \approx 0$ and $\Omega \lesssim 1$, and thus that the universe will expand forever.

What is finally needed to solve the system is the equation of state. Approximating the matter in the universe by a perfect fluid, the equation of state becomes

$$p = w\rho.$$

This implies that the cosmological constant behaves like a perfect fluid with $w = -1$. For matter with thermal energy substantially below the rest energy w is zero, while for ultrarelativistic matter, or massless particles, $w = 1/3$. Inserting these yields

$$a_\Lambda(t) \sim e^t \tag{3.13}$$

$$a_{\text{Matter}}(t) \sim t^{\frac{2}{3}} \tag{3.14}$$

$$a_{\text{Radiation}}(t) \sim t^{\frac{1}{2}}. \tag{3.15}$$

What entered here is an initial condition of $a = 0$.

To unify result, it is possible to combine different types of matter. This yields in terms of the fraction of the total density of each type in units of the critical density at a fixed time t_0

$$\frac{H^2}{H_0^2} = \frac{\Omega_{\text{Radiation}}}{a^4} + \frac{\Omega_{\text{Matter}}}{a^3} + \frac{\Omega_k}{a^2} + \Omega_\Lambda \tag{3.16}$$

where Ω_k is a suitable normalized version of k and H_0 is the value of H measured at the same fixed time t_0, usually today. This equation is, in principle, exactly solvable, given a very lengthy expression in terms of elliptic functions.

However, here just the important properties will be quoted for the observed values of the various Ω_i. These values are determined empirically. The values are $\Omega_k \approx 0$, $\Omega_{\text{Matter}} \approx 0.25$, $\Omega_{\text{Radiation}} \approx 0.01$, and $\Omega_\Lambda \approx 0.74$, where dark matter is included in matter. These values are the ones measured today. This modelling of the universe is also known as the cold-dark matter scenario with cosmological constant, briefly ΛCDM. Cold, because the dark matter is non-relativistic.

This yields that the universe started out from $a = 0$, i. e. a point, the so-called big bang. At that point also time started, and all worldlines originate from this event. There is no notion, within general relativity, of a before or outside. That is a very crucial insight. However, at the same time, this is a singularity, and thus likely an indication for a breakdown of gravity and the need for quantum gravity. This will be discussed in more detail in section **??**. For now, it is important there is a beginning, and no before. Afterwards it started to expand spatially.

After that, the universe was extremely hot, and thus radiation dominated it, yielding an expansion like (3.15). With the matter becoming more and more diluted, the temperature eventually dropped, and the equation became that of matter, slowing down the expansion to (3.14). However, eventually the cosmological constant will take over, yielding an exponentially accelerated expansion of the universe. This happens roughly around now. Thus, the ultimate fate of the universe is to become infinitely large, provided that nothing yet unknown kicks in.

Note that this is the expansion of the spatial hypersurface of the universe itself. This does not mean that matter will be blown apart. Because other interactions, especially attractive gravitational and electric ones, exist, matter will not tend to become homogeneously distributed, but clusters. Thus, even in the exponential late-time acceleration cluster of matter will stay together on galactic scales, but likely not on larger scales.

While this scenario is working well with a lot of evidence to be discuss in more detail in section **??**, it is not fully able to explain all observations. Especially the age of the oldest stars and globular clusters, about 13 billion years, is older than the age of the universe as determined by (3.16), which is about 10 billion years. The currently most likely solution is that of an additional effect making the universe much bigger in its very early stage, the so-called inflation. What essentially has happened is that at time-scales somewhere between temperature of order the Planck energy, $M_P \sim 10^{19}$ GeV and the so-called GUT scale of 10^{15} GeV the universe expanded exponentially by a factor of about e^{60}. While such an effect can be mediated by any first order phase transition of the matter in the universe, there is no known matter which will do it at such short time scales and by this amount. The standard model of particle physics will only yield some factor of e^{4-5} and far too late to be compatible with observations. This issue will be discussed in more detail in chapter **??**. For the moment, this will be just accepted as a feature, which essentially modifies the initial condition for the solution of (3.16). It pushes the age of the universe to about 14 billion years.

One of the decisive assumption in these derivations was the homogeneity and isotropy of the universe. That is, of course, violated on small scales. For the derivations to make sense, this is no problem, as on sufficiently large scales it is (approximately) true. Large scales need to be scales which are still substantially smaller than the size of the universe. The later is estimated to be of order 100 billion lightyears today, so more than an order of magnitude larger than the visible universe, which is of order 10 billion lightyears, the distances traversable by light in the age of the universe. This is a consequence of inflation that the actual universe is larger than the visible one.

The largest observed structure in the universe, supercluster of galaxies as well as voids and filaments, regions of low densities of galaxies and one-dimensional structure with high density of galaxies, are at least of order a few hundred million lightyears or more. Thus they approach a percent or more the size of the universe. With more observations incoming, this may worsen, eventually requiring to reproduce the evolution under less stringent assumptions on the structure of the universe. While this requires a numerical solution, anything where the fluctuations are not too extreme will still yield something approximately similar to the case presented here. Especially, any deviations can even influence the actual values of the Ω_i and H_0, as their definition is based on a homogeneous and isotropic universe.

3.5 Schwarzschild black hole and Kruskal coordinates

3.6 Kerr black hole

3.7 Gravitational waves

3.8 Wormholes, geons and other exotic solutions

3.9 ADM construction and momenta

Chapter 4

Canonical quantum gravity

4.1 Perturbation theory and its failure

4.1.1 Perturbative quantization

4.1.2 One-loop

4.1.3 Physical observables and augmenting perturbation theory

4.2 Beyond the Einstein-Hilbert Lagrangian

4.3 Functional renormalization group

4.4 Dynamical triangulation

4.5 Gravity as the square of Yang-Mills theory

Chapter 5

Beyond canonical quantum gravity

5.1 Loop-quantum gravity

5.2 Causal sets

5.3 Non-commutative geometry

5.4 String theory