Contents

2 Manifolds **21**

2.1 Motivation . . . 21

2.1.1 The failed Maxwell route . . . 21

2.1.2 The equivalence principles . . . 23

2.2 Manifold . . . 24

2.2.1 What is a manifold . . . 24

2.2.2 Coordinates and coordinate-independence . . . 25

2.3 Tensors and tangent spaces . . . 26

2.3.1 Vectors . . . 26

2.3.2 Covectors . . . 28

2.3.3 General tensors . . . 29

2.3.4 Symmetries . . . 31

2.4 The metric . . . 32

2.4.1 What is the metric . . . 32

2.4.2 Dot product . . . 33

2.4.3 Proper time and space intervals . . . 34

2.4.4 Integration on manifolds . . . 35

2.4.5 Levi–Civita tensor . . . 36

2.4.6 Tensor components and observations . . . 36

2.5 Back to the strong equivalence principle . . . 37

2.5.1 How close to local Minkowski space? . . . 37

2 Manifolds

2.1 Motivation

2.1.1 The failed Maxwell route

Let us start towards GR by setting out the problem and taking a plausible but wrong way to solve it. The question is how to reconcile special relativity and Newtonian gravity. In classical physics, Newton's gravitational force and Coulomb's electric force between two particles have the same form: both are proportional to the gravitational/electric charge of the particles and inversely proportional to the square of their distance r. Such interaction is not compatible with SR. In SR, signals cannot travel faster than the speed of light, but the $1/r^2$ force adjusts without delay when particles move, so information is conveyed with infinite speed.

In the case of the electric force, the contradiction is solved by the electromagnetic field that mediates the interaction. Charged particles do not affect each other directly, instead they interact with the electromagnetic field, which changes their motion. Since the response of the particles to the electromagnetic field (given by the Lorentz force) is local, as is the influence of the particles on the electromagnetic field (given by the Maxwell equations), and electromagnetic signals travel at the speed of light, information does not propagate faster than the speed of light. The $1/r^2$ Coulomb law is just an approximation in the limit of slow motion.

It thus seems sensible to try to solve the problem of the gravitational force in the same way, by promoting the gravitational field into a dynamical entity whose behaviour is locally determined by matter. In analogy with the electromagnetic case (where electromagnetic waves emerge from electrodynamics), we might expect this to introduce new phenomena.

In Newtonian physics, consider a system of particles labeled by n, with masses m (for simplicity all taken to be equal) and positions $x^i_{(n)}(t)$. Their motion under gravity is governed by Newton's second law with Newton's gravitational force,

$$m\frac{\mathrm{d}^2 x^i_{(n)}}{\mathrm{d}t^2} = F^i_{(n)} \quad = \quad -G_{\mathrm{N}} m \sum_{k\neq n} m \frac{x^i_{(n)} - x^i_{(k)}}{|\vec{x}_{(n)} - \vec{x}_{(k)}|^3} \ , \tag{2.1}$$

where $F^i_{(k)}$ is the gravitational force exerted on particle k by all other particles. In the case of a finite number of particles in infinite Euclidean space, the potential vanishes at infinity and the Newtonian force law (2.1) is equivalent to the Poisson equation

$$\nabla^2 \phi = 4\pi G_{\mathrm{N}} \rho_m \tag{2.2}$$

together with the force law[1]

$$m\frac{\mathrm{d}^2 x^i_{(n)}}{\mathrm{d}t^2} = -m\delta^{ij}\partial_j \phi(t,\vec{x})\Big|_{\vec{x}=\vec{x}_{(n)}} \ , \tag{2.3}$$

where $\nabla^2 \equiv \delta^{ij}\partial_i\partial_j$ is the Euclidean Laplacian, ρ_m is the mass density, and ϕ is the gravitational potential generated by all masses. (**Exercise.** Show this.)

This is nothing but a rewriting of the Newtonian force. However, this reformulation points to a way to generalise the Poisson constraint equation into a dynamical equation. A possibility that immediately suggests itself is to replace the Newtonian Laplacian with the relativistic D'Alembertian, $\delta^{ij}\partial_i\partial_j \to \eta^{\alpha\beta}\partial_\alpha\partial_\beta = -\frac{\partial^2}{\partial t^2} + \nabla^2$. This was proposed by the Finnish physicist Gunnar Nordström in 1912, along with a change to the equation of motion (2.1). Unfortunately the resulting equation is not covariant under Poincaré transformations, because the mass density is not invariant: mass is invariant, but the volume element changes due to Lorentz contraction. Nordström fixed this problem in 1913, introducing the first law of gravity that is consistent with SR:

$$\psi\Box\psi = -4\pi G_{\mathrm{N}} \eta^{\alpha\beta} T_{\alpha\beta} \ , \tag{2.4}$$

where ψ is related to the gravitational potential ($\psi = 1 + \phi$ recovers Newtonian gravity for small ϕ to leading order in ϕ), and $T_{\alpha\beta}$ is the **energy-momentum tensor**, which we will later discuss in detail; for now suffice to say that in the non-relativistic limit $T_{\alpha\beta} \simeq \delta_{\alpha 0}\delta_{\beta 0}\rho_m$. The equation (2.4) is a simple and elegant law of gravity consistent with SR. It is also wrong, meaning that it does not describe

[1] Even though the position of spatial indices makes no difference in Euclidean space with Cartesian coordinates, we prefer to match the positions of the indices on different sides of the equation, and hence write $\delta^{ij}\partial_j$ instead of ∂_i.

reality: relativistic gravity was not to be found by following the route of the Maxwell equations. Instead, Einstein and Hilbert, together with Michele Besso and Marcel Grossmann, took a different road, where gravity is described by the spacetime itself, instead of a field evolving in space. This required a radical reconceptualisation of spacetime. The resulting theory, GR, turned out not only to describe new phenomena in addition to the pull of gravity, it opened a whole new way of looking at physics geometrically.

2.1.2 The equivalence principles

In the development of GR, Einstein was guided by the curious property that objects fall at the same rate regardless of their composition (apart from the effect of air drag and other non-gravitational forces). Consider the equation of motion (2.3). The mass m has a different physical role on the left and on the right side of the equation. We can make this more transparent by using different symbols for the **inertial mass** m_{I} and the **gravitational mass** m_{G}, writing

$$m_{\mathrm{I}} \frac{\mathrm{d}^2 x^i}{\mathrm{d}t^2} = F^i = -m_{\mathrm{G}} \delta^{ij} \partial_j \phi \ . \tag{2.5}$$

Here m_{I} measures how much force has to be applied to the object to give it certain acceleration, i.e. inertia, whereas m_{G} determines how much force the field exerts on the object, i.e. gravity. In Newtonian mechanics these are independent quantities that have nothing to do with each other a priori. This can be appreciated by comparing to the electric force. The electric force is proportional to the charge q, so the resulting acceleration is proportional to q/m_{I}. It varies depending on the composition of the object. In contrast, it was known already at the end of the 19th century that the relative difference of the inertial and gravitational mass is less than 10^{-8}. Today this is known to hold to an accuracy of 10^{-15}. The property that all objects move the same way under gravity is known as the **weak equivalence principle**.

This suggests that inertia and gravity are intrinsically related. As all objects fall the same way, an observer falling freely under a constant gravitational force cannot by any local observation distinguish that they are in a gravitational field. If you are inside an elevator and the cord is cut, you won't feel the pull of the Earth, but will rather be weightless as in space. Einstein called the realisation in 1907 that a falling person does not feel their weight "the happiest thought of my life". Newton's first law erased the distinction between staying at rest and moving at constant velocity, and GR unifies moving at constant velocity and falling freely in a gravitational field, intertwining inertia and gravity.

If the gravitational field is not constant in space and time, it is possible to distinguish the effect of gravity because objects at different points of spacetime fall differently. If the elevator is large enough, people on the left side will move towards a slightly different direction than people on the right, as all are pulled towards the centre of the Earth. Similarly, the motion will be different at different times, as the distance to the Earth changes. Gravity is exactly absent pointwise, and approximately over a spacetime region that is small compared to the scale over

which the gravitational field varies significantly. This suggests that gravity is tied to the relations between spacetime points. The property that locally laws of physics reduce to non-gravitational laws (in GR to those of SR) is known as the **strong equivalence principle**.[2]

The weak and the strong equivalence principle were used as guidelines when searching for GR. Such principles are scaffolding that can be cast aside after the full mathematical formulation of a physical theory has been discovered. They are not extra structures that need to be imposed on the theory, but rather results that may follow from the theory to varying extent. In the case of GR, the mathematical structure of the theory will give us a precise definition of the weak and the strong equivalence principle, and show when they are satisfied and whether they can be violated.

2.2 Manifold

2.2.1 What is a manifold

Let us move from the motivation sketched above to mathematical formulation. We will be more careful with notation and definition than in chapter 1. If we want to implement the idea that gravity is a property of spacetime and that it is locally absent, we need a description of spacetime that looks like Minkowski space pointwise. A **manifold** provides such a mathematical structure. A d-dimensional differentiable manifold M is defined as a topological space that satisfies the following conditions[3]:

1. There exists a family of pairs $\{(U_n, \varphi_n)\}$, where U_n is an open set on M, and $\varphi_n : U_n \to \mathbb{R}^d$ is a homeomorphic (i.e. bijective, continuous, and having a continuous inverse) map from U_n to $\mathbb{R}^d$.

2. The image $\varphi_n(U_n)$ is an open set and $\cup_n U_n = M$.

3. For U_n and U_m such that $U_n \cap U_m \neq \emptyset$, the map $\varphi_n \circ \varphi_m^{-1}$ from $\varphi_m(U_n \cap U_m)$ to $\varphi_n(U_n \cap U_m)$ is C^∞.

The pair (U_n, φ_n) is called a **chart** or a **map**, and the set of all charts $\{(U_n, \varphi_n)\}$ is called an **atlas**. A map is also called a **coordinate system**, and φ_n are called **coordinates**.

Consider two maps, $\varphi : U \to \mathbb{R}^d$ and $\varphi' : U' \to \mathbb{R}^d$, which assign the numbers (also called coordinates) x^α and x'^α, respectively, to points that lie in $U \cap U'$ (assumed to be non-empty). See figure 1.

The first and the third property above imply that in $U \cap U'$ the coordinates are functions of each other,

$$\begin{cases} x^\alpha = x^\alpha(x'^0, \ldots, x'^{d-1}) \equiv x^\alpha(x') \\ x'^\alpha = x'^\alpha(x^0, \ldots, x^{d-1}) \equiv x'^\alpha(x) \ , \end{cases} \tag{2.6}$$

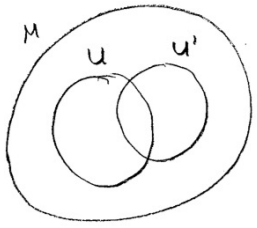

Figure 1: Two overlapping sets U and U'.

[2] There are slightly different definitions of the strong equivalence principle, and a third principle called the Einstein equivalence principle is also sometimes introduced. For example, see Clifford Will's review paper https://arxiv.org/abs/1403.7377.

[3] We only consider real manifolds, relevant for GR. For complex manifolds, $\mathbb{R}^n$ is replaced by $\mathbb{C}^n$.

and that all derivatives of these functions exist and are continuous. This implies that the Jacobian matrices

$$M^{\alpha}{}_{\beta} \equiv \frac{\partial x'^{\alpha}}{\partial x^{\beta}} \ , \qquad (M^{-1})^{\alpha}{}_{\beta} \equiv \frac{\partial x^{\alpha}}{\partial x'^{\beta}} \tag{2.7}$$

are

$$\text{nonsingular}: \quad \det(M^{\alpha}{}_{\beta}) \neq 0 \ , \quad \text{and} \tag{2.8}$$

$$\text{the inverse of each other:} \quad \frac{\partial x'^{\alpha}}{\partial x^{\gamma}} \frac{\partial x^{\gamma}}{\partial x'^{\beta}} = \delta^{\alpha}{}_{\beta} \ . \tag{2.9}$$

The chain rule of differentiation reads

$$\frac{\partial}{\partial x^{\alpha}} = \frac{\partial x'^{\beta}}{\partial x^{\alpha}} \frac{\partial}{\partial x'^{\beta}} = M^{\beta}{}_{\alpha} \frac{\partial}{\partial x'^{\beta}} \ . \tag{2.10}$$

These relations hold only in the overlap region $U \cap U'$. The question of how the parts of U that do not overlap with U' look like in the coordinates x^{α} is meaningless. Note that while we have the topological notion of a coordinate neighbourhood on the manifold, we have not yet introduced any geometry, so distances and straight lines are not defined. For that we need more structure, which will be given by a generalisation of the Minkowski metric, to be defined later.

2.2.2 Coordinates and coordinate-independence

The charts (U_n, φ_n) provide a formal definition of coordinate systems, i.e. ways of assigning real numbers x^{α} to spacetime points, which we discussed loosely in chapter 1. For a general manifold, it is not possible to assign these numbers in such a way that all points in each other's infinitesimal neighbourhood have coordinate values that are infinitesimally close, and the entire manifold cannot in general be covered by a single map (U_n, φ_n). We will encounter a dramatic example of this when we discuss black holes, but a simple case is given by the unit two-sphere S^2. If we use spherical coordinates, the angular coordinate φ jumps from 2π to 0 when crossing the prime meridian. This can be avoided by extending the φ coordinate to the entire real axis, but then points no longer have unique coordinates, in contradiction with what is assumed about charts in the definition of a manifold. Also, in either case, the coordinate φ does not have a unique value at the north pole nor at the south pole. We need two copies of $\mathbb{R}^2$ to cover S^2 so that all coordinate values of neighbouring points are infinitesimally close, and every point is assigned a unique number on each chart.

The two-sphere example also shows that, while the local topology of a manifold is by definition the same as in $\mathbb{R}^n$, the global topology can be different. Another example would be a two-torus T^2, i.e. a rectangular piece of $\mathbb{R}^2$ with opposite ends of the rectangle identified without twist. Embedded into $\mathbb{R}^3$, a two-torus looks like a doughnut. The geometry is not only locally but globally $\mathbb{R}^2$, but due to non-trivial topology, the manifold cannot be covered with a single chart that would assign unique coordinates and give neighbouring points close coordinate values. Note that while it is simple and often useful to consider two-dimensional surfaces embedded in

$\mathbb{R}^3$ as examples of manifolds, in general manifolds are not considered to embedded in a higher-dimensional space, and their properties are intrinsic, not defined by any embedding.

The picture is that physics lives on the manifold, and we describe parts of it using coordinates. Coordinate systems by themselves have no physical meaning: no coordinate system is preferred over another. This means that the physics has to be invariant (or covariant, i.e. equations have to retain their their shape) under coordinate transformations. This requirement is called **diffeomorphism invariance**. It imposes strong restrictions on how physical quantities are described. (This is analogous to the way rotation, translation and Galilei symmetry constrain the possible laws of physics in the Newtonian case, but diffeomorphism invariance is not associated with symmetries of the spacetime.) This means that all physical quantities are described in terms of **tensors**, to which we now turn.

2.3 Tensors and tangent spaces

2.3.1 Vectors

Let us first introduce **vectors** on the manifold. For vectors in general, the commutative operations of addition and multiplication by scalars are defined and satisfy the following properties:

$$(a+b)(\underline{A}+\underline{B}) = a\underline{A} + b\underline{B} + a\underline{A} + b\underline{B} \ , \tag{2.11}$$

where a, b are real numbers and $\underline{A}$, $\underline{B}$ are vectors.

In GR, a vector $\underline{A}$ is a particular type of coordinate-independent map (which we define below) from functions on the manifold to functions on the manifold. A **vector field**, denoted $\underline{A}(x)$, is a rule that attaches exactly one vector $\underline{A}$ to each point x on the manifold (or a subset of the manifold).

Let F be the space of all smooth functions from the manifold to the real numbers, i.e. all C^∞ maps $f : M \to \mathbb{R}$. Take now a parametrised smooth curve on the manifold, i.e. a C^∞ map $\gamma : \mathbb{R} \to M$ from the real numbers to the manifold. Let λ be the parameter that determines the position on the curve. Given a curve γ, the parameter λ determines a point on the manifold, so we can consider a function f as a function of λ: $f = f[x(\lambda)]$. Taking a derivative with respect to λ then maps F to itself: $f \to \frac{\mathrm{d}f}{\mathrm{d}\lambda}$. For a curve γ passing through point p, we have $f\big|_p \to \frac{\mathrm{d}f}{\mathrm{d}\lambda}\big|_p$. This is a directional derivative of f at point p, and the corresponding derivative operator is denoted by $\frac{\mathrm{d}}{\mathrm{d}\lambda}\big|_p$. The **tangent space** T_p at point p is the set of all directional derivative operators along all possible parametrised curves that pass through p (you can check that this set forms a vector space), illustrated in figure 2. The elements of T_p are vectors.

Vectors are written in terms of **components** A^α and **basis vectors** $\underline{e}_\alpha$ as

$$\underline{A} = A^\alpha \underline{e}_\alpha \ . \tag{2.12}$$

Note that the subscript α in $\underline{e}_\alpha$ does not label components, but basis vectors. The components of the basis vectors are $(\underline{e}_\alpha)^\beta$. Vectors are independent of the choice of

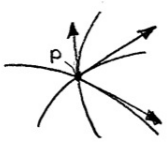

Figure 2: Curves passing through p, with tangent vectors.

coordinates, but both the components and the basis vectors in general depend on the choice of coordinates.

There are different ways of choosing the basis vectors for the tangent space. In this course, we will mostly use the **coordinate basis**, defined as follows. Choose a coordinate system and a point p. On a d-dimensional manifold, we use the coordinates x^α to define d parametrised curves through p by the following procedure. Choose one coordinate labeled β, and take the curve where x^α is constant for every value of α other than β, as illustrated in figure 3. Each such curve defines a map $f \to \left.\frac{\partial f}{\partial x^\beta}\right|_p$ along the line on which x^α = constant for all $\alpha \neq \beta$. Therefore the partial derivatives ∂_β at p are elements of T_p.

Using the chain rule, we can write $\frac{\mathrm{d}f}{\mathrm{d}\lambda}$ for any curve as

$$\frac{\mathrm{d}f}{\mathrm{d}\lambda} = \frac{\partial f}{\partial x^\alpha}\frac{\mathrm{d}x^\alpha}{\mathrm{d}\lambda} \quad \Leftrightarrow \quad \frac{\mathrm{d}}{\mathrm{d}\lambda} = \frac{\mathrm{d}x^\alpha}{\mathrm{d}\lambda}\partial_\alpha \ , \tag{2.13}$$

so the set of partial derivatives $\{\partial_\alpha\}$ forms a basis of T_p. Since there are d partial derivatives ∂_α, we have dim T_p = dim M. So the tangent spaces have the same dimension as the manifold, and d independent basis vectors. The basis vectors $\underline{e}_\alpha = \partial_\alpha$ are called the **coordinate basis vectors**.[4] The tangent space T_p has the same geometry as Minkowski spacetime.

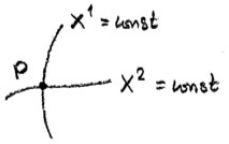

Figure 3: Constant coordinate lines as parametrised curves through p.

By definition, the components of the coordinate basis vectors are

$$(\underline{e}_\alpha)^\beta = (\partial_\alpha)^\beta = \delta_\alpha{}^\beta \ . \tag{2.14}$$

Under a change of coordinates, according to the chain rule the coordinate basis vectors (not their components!) transform as

$$\partial_\alpha = \frac{\partial}{\partial x^\alpha} \to \partial'_\alpha = \frac{\partial}{\partial x'^\alpha} = \frac{\partial x^\beta}{\partial x'^\alpha}\frac{\partial}{\partial x^\beta} = (M^{-1})^\beta{}_\alpha \partial_\beta \ . \tag{2.15}$$

In the coordinate basis, a general vector reads $\underline{A} = A^\alpha \partial_\alpha$. Because a vector is by definition independent of coordinates, its components transform inversely to the basis vectors,

$$\underline{A} = A^\alpha \partial_\alpha = A'^\alpha \partial'_\alpha \ . \tag{2.16}$$

[4] Once we define a dot product, we will find that in general the coordinate basis vectors ∂_α are not normalised to unity, nor orthogonal to each other. On a curved manifold, it is always possible to find a coordinate system such that the coordinate basis vectors are orthonormal at a given point p, but in general this cannot be extended to the neighbourhood of the point.

As the basis vectors transform with $(M^{-1})^{\beta}{}_{\alpha}$, the components transform with $M^{\alpha}{}_{\beta}$.

$$A^{\alpha} \to A'^{\alpha} = \frac{\partial x'^{\alpha}}{\partial x^{\beta}} A^{\beta} = M^{\alpha}{}_{\beta} A^{\beta} \ . \tag{2.17}$$

Lorentz transformations are the special case when the coordinate transformation is linear, and the Jacobian matrix is constant. In SR, Lorentz transformations map between coordinate systems moving at constant velocity, and the constancy of the transformation matrix expresses a symmetry of the theory, namely the homogeneity of spacetime. In GR, the transformations map between arbitrary coordinate systems, and they are not related to symmetries of the spacetime. (In GR, spacetimes in general have no symmetry.)

An example may clarify the role of vector fields as linear differential operators that map functions to functions. Consider a curve parametrised by λ. The vector field $\underline{u} = \frac{\mathrm{d}}{\mathrm{d}\lambda}$ gives at every point along the curve the vector that is tangent to the curve, i.e. the tangent vector. Acting on a function f, it gives

$$\underline{u}(f) = \frac{\mathrm{d}f}{\mathrm{d}\lambda} = \frac{\mathrm{d}x^{\alpha}}{\mathrm{d}\lambda} \partial_{\alpha} f = u^{\alpha} \partial_{\alpha} f \ , \tag{2.18}$$

which gives the observed rate of change of f in the direction of the curve. If the curve is timelike (i.e. a worldline), we can choose $\lambda = \tau$, where τ is the proper time, and $\underline{u}$ then is the velocity of an observer moving on the worldline. (This is a bit premature: we will soon in sections 2.4.2 and 2.4.3 define the terms timelike and proper time.)

A vector field maps functions to functions. We define the **commutator** of vector fields $\underline{A}$ and $\underline{B}$ as

$$[\underline{A}, \underline{B}](f) \equiv \underline{A}(\underline{B}(f)) - \underline{B}(\underline{A}(f)) = A^{\beta} \partial_{\beta} (B^{\alpha} \partial_{\alpha} f) - B^{\beta} \partial_{\beta} (A^{\alpha} \partial_{\alpha} f) \ . \tag{2.19}$$

This commutator is also known as the **Lie derivative** of the vector field $\underline{B}$ along the vector field $\underline{A}$, $\mathcal{L}_{\underline{A}} \underline{B} \equiv [\underline{A}, \underline{B}]$. It is easy to show that the commutator $[\underline{A}, \underline{B}]$ is also a vector field, whereas the operator $\underline{A} \circ \underline{B}$ defined by $f \to \underline{A}(\underline{B}(f))$ is in general not a vector field. (**Exercise.** Show this.)

2.3.2 Covectors

Now that we have defined vectors and tangent spaces on a manifold, we can define **tensors** of all kinds. Let us begin with **dual vectors**, also known as **covectors** and **one-forms**. To each tangent space T_p, we associate the **cotangent space** T_p^*, defined as the set of all linear maps $\tilde{\omega} : T_p \to \mathbb{R}$ from the tangent space to the real numbers. The linear maps are called dual vectors. Because of linearity, we have

$$\tilde{\omega}(a\underline{A} + b\underline{B}) = a\tilde{\omega}(\underline{A}) + b\tilde{\omega}(\underline{B}) \ , \tag{2.20}$$

where a, b are real numbers. Therefore,

$$\tilde{\omega}(\underline{A}) = \tilde{\omega}(A^{\alpha} \underline{e}_{\alpha}) = A^{\alpha} \tilde{\omega}(\underline{e}_{\alpha}) \equiv A^{\alpha} \omega_{\alpha} \ , \tag{2.21}$$

where we have defined $\tilde{\omega}(\underline{e}_\alpha) \equiv \omega_\alpha$. In terms of the **dual basis vectors** $\tilde{e}^\alpha$, a dual vector reads

$$\tilde{\omega} = \omega_\alpha \tilde{e}^\alpha \ . \tag{2.22}$$

Like vectors, the dual vectors are defined on the manifold (more precisely, in the tangent spaces), and are independent of the coordinate system. Under a change of coordinates, the components A^α transform as $A^\alpha \to M^\alpha{}_\beta A^\beta$. As the map (2.21) is invariant under coordinate transformations, the quantities ω_α transform as $\omega_\alpha \to (M^{-1})^\beta{}_\alpha \omega_\beta$. From (2.22) we then see that the dual basis vectors (not their components!) transform as $\tilde{e}^\alpha \to M^\alpha{}_\beta \tilde{e}^\beta$. The transformation rule is easy to remember: all upstairs labels transform with M and all downstairs labels transform with M^{-1}, whether we consider components or basis vectors.

We choose the dual basis vectors so that

$$\tilde{e}^\alpha(\underline{e}_\beta) = \delta^\alpha{}_\beta \ . \tag{2.23}$$

We have $\dim T_p^* = \dim T_p$. Taking the dual of the dual vector space gives back the vector space.

In a coordinate basis, partial derivatives form the basis of the vector space. The basis vectors $\tilde{e}^\alpha$ of covector space are then often denoted by $\tilde{\mathrm{d}}x^\alpha$, or simply by $\mathrm{d}x^\alpha$, similar to coordinate differentials. We will not use this notation.

2.3.3 General tensors

A covector is a rule that gives you a number if you give it a vector. We can also read this in the other direction, by keeping the vector fixed and varying the covector. We then have a rule that gives a number if you give it a covector. So we can also view vectors as linear maps from the cotangent space to the real numbers, $\underline{A} : T_p^* \to \mathbb{R}$. Vectors and covectors are thus in a completely symmetric position.

Vectors and covectors are examples of tensors. We now define a general tensor of type (r, s) as a linear map from r copies of the cotangent space and s copies of the tangent space to the real numbers:

$$T : \underbrace{T_p^* \times \cdots \times T_p^*}_{r \text{ times}} \times \underbrace{T_p \times \cdots \times T_p}_{s \text{ times}} \to \mathbb{R} \ . \tag{2.24}$$

Like vectors and covectors, all tensors are coordinate-independent, and are defined on the manifold. A tensor can be viewed as a slot machine with r round and s square slots, which gives you a real number once you insert the requisite amount of round and square tokens. The **order** (also called **rank** and **degree**) of a tensor is $r+s$: it is the total number of slots. Vectors are type $(1, 0)$ tensors and covectors are type $(0, 1)$ tensors. A type $(0, 0)$ tensor is defined to be just a single number. Just as we defined vector fields, we define a general **tensor field** as a rule that attaches a tensor to every point on the manifold (or a subset of the manifold). Type $(0, 0)$ tensor fields are scalar functions, which attach a single number to every point on the manifold (or its subset). Tensors of type $(0, s)$ (with $s \geq 1$) are called **covariant**, tensors of type $(r, 0)$ (with $r \geq 1$) are called **contravariant**, and other tensors of

rank ≥ 2 are called **mixed**. Correspondingly, component indices that are down are called covariant and those that are up are called contravariant.

Consider a rank 2 covariant tensor $T : T_p \times T_p \to \mathbb{R}$. It is a rule that gives a number given two vectors, $T(\underline{A}, \underline{B}) \in \mathbb{R}$. If we give only one vector, we obtain a rule that gives a number if given a vector, i.e. we get a covector: $\tilde{\omega} \equiv T(\underline{A}, \ ) : T_p \to \mathbb{R}$. Hence we can view T as a rule that attaches a covector to a vector, $T : T_p \to T_p^*$. In fact, we get two such rules, depending on whether we insert the vector in the first or second slot: $T(\ , \underline{A})$ is also a covector, and it is different from $T(\underline{A}, \ )$ unless T is symmetric (we discuss symmetry in section 2.3.4 below). Similarly, a rank 2 contravariant tensor is two maps $T_p^* \to T_p$, and a rank 2 mixed tensor is a map $T_p \to T_p$ and a map $T_p^* \to T_p^*$. This generalises in an obvious way to tensors of higher rank.

In terms of components, a type (r, s) tensor reads (the symbol $\otimes$ denotes tensor product, defined below; the symbol is often omitted)

$$T = T^{\alpha_1 \ldots \alpha_r}{}_{\beta_1 \ldots \beta_s} \underline{e}_{\alpha_1} \otimes \ldots \otimes \underline{e}_{\alpha_r} \otimes \tilde{e}^{\beta_1} \otimes \ldots \otimes \tilde{e}^{\beta_s} \ . \tag{2.25}$$

Under a change of coordinates, the components transform as

$$T^{\alpha_1 \ldots \alpha_r}{}_{\beta_1 \ldots \beta_s} \to M^{\alpha_1}{}_{\mu_1} \ldots M^{\alpha_r}{}_{\mu_r} (M^{-1})^{\nu_1}{}_{\beta_1} \ldots (M^{-1})^{\nu_s}{}_{\beta_s} T^{\mu_1 \ldots \mu_r}{}_{\nu_1 \ldots \nu_s} \ . \tag{2.26}$$

If we are given a set of functions and want to know whether they are the components of a tensor, the recipe is easy. If they transform according to the rule (2.26), they are the components of a tensor in the coordinate basis. If not, they are not such components.[5] Consider the partial derivative of a scalar function (recall the transformation rule for partial derivatives in (2.15))

$$\partial_\alpha f \to (M^{-1})^\beta{}_\alpha \partial_\beta f \ . \tag{2.27}$$

Comparing to (2.26), we see that the gradient is a type $(0, 1)$ tensor, i.e. a covector. Consider now the partial derivative of the components of a covector field

$$\begin{aligned} \partial_\alpha A_\beta &\to (M^{-1})^\gamma{}_\alpha \partial_\gamma [(M^{-1})^\delta{}_\beta A_\delta] \\ &= (M^{-1})^\gamma{}_\alpha (M^{-1})^\delta{}_\beta \partial_\gamma A_\delta + (M^{-1})^\gamma{}_\alpha \partial_\gamma (M^{-1})^\delta{}_\beta A_\delta \ . \end{aligned} \tag{2.28}$$

The first term is the tensor transformation law, but the second term is an extra piece. If (and only if) we restrict to coordinate transformations where the Jacobian matrix is constant, the partial derivatives of tensor components are the components of a tensor. In particular, this is true for the symmetry transformations that leave the Minkowski metric invariant (i.e. Poincaré transformations). Otherwise partial derivatives of the components of tensors (including second and higher derivatives of scalars) are not the components of any tensor. They make sense only in a coordinate system, not on the manifold. In the next chapter we will introduce a derivative that is defined on the manifold as an operation on tensors and not just on components, the **covariant derivative**. (In this chapter, we have already encountered another derivative defined on the manifold, the Lie derivative.)

[5] If we are given the right number of functions in a certain coordinate system, we can always define a tensor by assuming that they are the components and transform according to (2.26).

As the components completely define a tensor (when the basis is fixed), we will often lazily refer to the components as the tensor, but we should understand the distinction. The useful thing about writing the components is that they immediately identify the type of the tensor. For example, if we write $T_{\alpha\beta}$, we see that we are talking about a tensor of type $(0,2)$, which the notation A doesn't tell us, unless we expand in terms of the basis covectors, $T = T_{\alpha\beta}\tilde{e}^{\alpha}\tilde{e}^{\beta}$.

Multiplication of tensors with each other is defined with the **tensor product**, denoted by $\otimes$. It maps a tensor of type (r_1, s_1) and a tensor of type (r_2, s_2) into a tensor of type $(r_1 + r_2, s_1 + s_2)$. It can be defined in terms of the components simply by saying that the components of $A \otimes B$ are the components of A multiplied by the components of B, with the set of indices kept in the same order. For example, for tensors of type $(2,0)$ and $(0,3)$ we have

$$(A \otimes B)^{\alpha\beta}{}_{\gamma\delta\epsilon} = A^{\alpha\beta}B_{\gamma\delta\epsilon} \ . \tag{2.29}$$

The tensor product is non-commutative, in general $A \otimes B \neq B \otimes A$. For example,

$$(B \otimes A)_{\alpha\beta\gamma}{}^{\delta\epsilon} = B_{\alpha\beta\gamma}A^{\delta\epsilon} \ . \tag{2.30}$$

2.3.4 Symmetries

Tensors can be **symmetric** or **antisymmetric**. For example, a covariant rank 2 tensor is symmetric if $T(\underline{A}, \underline{B}) = T(\underline{B}, \underline{A})$. In terms of components, symmetry is denoted by parentheses around the indices: for a symmetric tensor we have $T_{\alpha\beta} = T_{(\alpha\beta)}$, where

$$T_{(\alpha\beta)} \equiv \frac{1}{2}(T_{\alpha\beta} + T_{\beta\alpha}) \ . \tag{2.31}$$

Similarly, a covariant rank 2 tensor is antisymmetric if $T(\underline{A}, \underline{B}) = -T(\underline{B}, \underline{A})$. Antisymmetry is denoted by square brackets around the indices. For a rank 2 tensor we have

$$T_{[\alpha\beta]} \equiv \frac{1}{2}(T_{\alpha\beta} - T_{\beta\alpha}) \ . \tag{2.32}$$

An arbitrary rank 2 tensor can be decomposed into the symmetric and the antisymmetric part as $T_{\alpha\beta} = T_{(\alpha\beta)} + T_{[\alpha\beta]}$.

For a rank n tensor, the rules for symmetrisation and antisymmetrisation are:

$$\begin{aligned} T_{(\alpha_1 \ldots \alpha_n)} &\equiv \frac{1}{n!}(\text{sum over all index permutations}) \\ T_{[\alpha_1 \ldots \alpha_n]} &\equiv \frac{1}{n!}(\text{alternating sum over all index permutations}) \ , \end{aligned} \tag{2.33}$$

where alternating sum means that every term comes with a factor of $(-1)^p$, where p is the number of indices permuted, so odd permutations have a minus sign. For example,

$$T_{[\alpha\beta\gamma]} = \frac{1}{6}(T_{\alpha\beta\gamma} - T_{\beta\alpha\gamma} + T_{\beta\gamma\alpha} - T_{\alpha\gamma\beta} + T_{\gamma\alpha\beta} - T_{\gamma\beta\alpha}) \ . \tag{2.34}$$

Up indices are only ever symmetrised or antisymmetrised with each other, and the same for down indices. An index that is not included in the symmetrisation or antisymmetrisation is enclosed inside the vertical bars ||. For example,

$$T_{(\alpha|\beta|\gamma)} = \frac{1}{2}(T_{\alpha\beta\gamma} + T_{\gamma\beta\alpha}) \ . \tag{2.35}$$

2.4 The metric

2.4.1 What is the metric

So far, we have not imposed any geometry on the manifold. Let's do that now. In GR, gravity is an aspect of the geometry of spacetime, encoded in the **metric** g, and we write the spacetime as the pair (M, g), where M is the manifold. The metric is a symmetric type $(0, 2)$ tensor field whose components $g_{\alpha\beta}(x) = g_{\beta\alpha}(x)$ form a non-singular matrix, i.e. $\det(g_{\alpha\beta}) \neq 0$. This implies that the inverse matrix exists. It is denoted by $g^{\alpha\beta}$, so by definition $g^{\alpha\gamma}g_{\gamma\beta} = \delta^{\alpha}{}_{\beta}$. The quantities $g^{\alpha\beta}$ are the components of a type $(2, 0)$ tensor, called the inverse metric. Note that while we still use the symbol $\eta_{\alpha\beta} = \mathrm{diag}(-1, 1, 1, 1)$, it does not denote the components of a tensor except in Minkowski space.

We can always find a coordinate system where the metric at a point is constant and diagonal, and the diagonal values are ± 1 (we will show this below in section 2.5). A manifold is then categorised according to the signs of the diagonal entries.[6] This set of ± 1 values is called the **signature** of the metric. If they are all positive, the manifold is **Riemannian**, i.e. it is a space. If also the metric is everywhere constant, we have a Euclidean manifold. If some of the entries are negative, the manifold is **pseudo-Riemannian**. In the case when exactly one entry is negative and there are at least two dimensions, the manifold is **Lorentzian**, i.e. it is a spacetime. If also the metric is everywhere constant, the manifold is Minkowki space. We will concentrate on four-dimensional Lorentzian manifolds.

The metric defines a number of key properties of spacetime, some of which are simple generalisations of the properties of Minkowski space, while others are new. The metric gives us the following.

1) The notion of past and future, the notion of timelike, spacelike and null intervals, and the notion of causality.

2) Proper length, proper time interval, proper area, proper volume and proper four-volume. (Proper means physical, i.e. coordinate-independent.)

3) The generalisation of the Newtonian gravitational potential as the field that describes gravity.

4) The definition of locally inertial non-rotating frames, a generalisation of the inertial coordinates used in Newtonian physics and SR.

[6] Physics does not change if we change the overall sign of the metric; this a matter of convention. So to be precise, instead of all entries being positive, we should say they all have same sign, and instead of exactly one entry being negative, we should say that exactly one entry has the opposite sign than the others.

5) A bijective map from the tangent space to the cotangent space, and the inner product of vectors.

If we identify straight curves with extremal curves (those that give a local minimum or maximum of the distance), the metric also gives us a sixth property:

6) The notion of a straight path in spacetime.

In GR, this identification is made, but distance and direction are in general independent concepts, so other choices are possible, leading to theories of spacetime different from GR (or sometimes just GR in a different guise: we will discuss this in section 6.1.5, in the second part of the course).

As as in SR, we use the metric to "raise and lower indices", i.e. to map a tensor of type (r, s) to a tensor of type $(r-1, s+1)$ or $(r+1, s-1)$. As we earlier noted, any type $(0, 2)$ tensor gives a map from the tangent space to the cotangent space. We identify the metric as the preferred map, and consider the tensors obtained by mapping with it to be versions of the same tensor. For example, we identify each vector with exactly one covector. We refer to the components of such tensors as different kinds (covariant, contravariant, mixed) components of the same tensor. Consider a type $(0, 2)$ tensor with components $T_{\alpha\beta}$. We define

$$\begin{aligned} T^{\alpha}{}_{\beta} &\equiv g^{\alpha\gamma} T_{\gamma\beta} \\ T_{\alpha}{}^{\beta} &\equiv g^{\beta\gamma} T_{\alpha\gamma} \\ T^{\alpha\beta} &\equiv g^{\alpha\gamma} g^{\beta\delta} T_{\gamma\delta} \ . \end{aligned} \tag{2.36}$$

The above operations are called **contracting** with the metric. In general, the **contraction** of two tensors is an operation that produces a type $(r_1 + r_2 - 1, s_1 + s_2 - 1)$ tensor out of a type (r_1, s_1) tensor and a type (r_2, s_2) tensor. Let us give the rule in the case of a type $(2, 0)$ tensor A and a type $(0, 3)$ tensor B; the generalisation to arbitrary tensors is straightforward. If we choose to contract the first and the second index, the components of the contracted tensor C are

$$C^{\alpha}{}_{\beta\gamma} = A^{\alpha\delta} B_{\beta\delta\gamma} = A^{\alpha}{}_{\delta} B_{\beta}{}^{\delta}{}_{\gamma} \tag{2.37}$$

We could contract any other two indices, resulting in a different tensor.

2.4.2 Dot product

A particular case of contraction is when we contract a tensor of type $(1, 0)$ and a tensor of type $(0, 1)$, i.e. a vector and a covector. The result is a scalar. We can also read the operation as first mapping the covector into a vector, and then contracting both of them with the metric. This gives the **dot product** (also called the **scalar product** and the **inner product**) of two vectors, denoted by

$$\underline{A} \cdot \underline{B} \equiv g(\underline{A}, \underline{B}) = g_{\alpha\beta} A^{\alpha} B^{\beta} = g^{\alpha\beta} A_{\alpha} B_{\beta} = A_{\alpha} B^{\alpha} = A^{\alpha} B_{\alpha} \ . \tag{2.38}$$

If the dot product of two vectors is zero, they are **orthogonal:**

$$\underline{A} \cdot \underline{B} = 0 \quad \Leftrightarrow \quad \underline{A} \perp \underline{B} \ . \tag{2.39}$$

Note that because the metric is not positive-definite, a vector can be orthogonal to itself without being the zero vector.

We categorise vectors into spacelike, null and timelike as follows:

$$\begin{cases} \text{spacelike} & g(\underline{A},\underline{A}) > 0 \\ \text{lightlike (null)} \quad \text{if} & g(\underline{A},\underline{A}) = 0 \\ \text{timelike} & g(\underline{A},\underline{A}) < 0 \ . \end{cases} \tag{2.40}$$

If the time direction is defined by the timelike four-vector $\underline{u}$, we call a timelike vector $\underline{A}$ future timelike (or future-oriented) if $g(\underline{u},\underline{A}) < 0$ and past timelike (or past-oriented) if $g(\underline{u},\underline{A}) > 0$. Future and past null vectors are defined with $\underline{u}$ in the same way. Curves on the manifold are denoted the same way, according to whether their tangent vector is spacelike, null or timelike, and (in the case of timelike and null vectors) future- or past-oriented.

A vector field can in principle vary between being spacelike, null and timelike depending on location, but many physically relevant vector fields are everywhere timelike, null or spacelike; likewise for future- and past-oriented.

2.4.3 Proper time and space intervals

The spacetime distance between a pair of infinitesimally nearby points with coordinates x^α and $x^\alpha + \mathrm{d}x^\alpha$ is

$$\mathrm{d}s = \sqrt{|g_{\alpha\beta}\mathrm{d}x^\alpha \mathrm{d}x^\beta|} \ , \tag{2.41}$$

where the distance is a length or a proper time (or zero) depending on whether the quantity inside the absolute value sign is positive or negative (or zero). This distance is the square root of the absolute value of **the line element**

$$\mathrm{d}s^2 = g_{\alpha\beta}\mathrm{d}x^\alpha \mathrm{d}x^\beta \ . \tag{2.42}$$

Along a spacelike curve $x^\alpha(\lambda)$, the **proper length** from point p to point q is defined as

$$L = \int_p^q \mathrm{d}\lambda \sqrt{g_{\alpha\beta}\frac{\mathrm{d}x^\beta}{\mathrm{d}\lambda}\frac{\mathrm{d}x^\beta}{\mathrm{d}\lambda}} \ . \tag{2.43}$$

Along a timelike curve $x^\alpha(\lambda)$, the **proper time interval** from p to q is defined as

$$\tau = \int_p^q \mathrm{d}\lambda \sqrt{-g_{\alpha\beta}\frac{\mathrm{d}x^\alpha}{\mathrm{d}\lambda}\frac{\mathrm{d}x^\beta}{\mathrm{d}\lambda}} \tag{2.44}$$

A null curve has zero length. Note that distances between two points depend on the curve along which the distance is measured. We can then ask what kind of a curve gives the smallest or the largest distance between two points. Such extremal curves are important in GR, and we will consider them in chapter 3.

2.4.4 Integration on manifolds

Let us discuss how to integrate quantities on a four-dimensional manifold. If we take the components of a vector field $A^\alpha(x)$ and simply integrate over some coordinate four-volume with the integration measure $\mathrm{d}^4x = \mathrm{d}x^0\mathrm{d}x^1\mathrm{d}x^2\mathrm{d}x^3$, the resulting quantities will not be the components of a vector – in other words, this operation is defined on the coordinate patch, but not on the manifold. There are two reasons.

The first problem is that the quantities $A^\alpha(x)$ and $A^\alpha(x')$ with $x \neq x'$ belong to different vector spaces. The first quantity is the components of a vector in the tangent space at x and the second is the components of a vector in the tangent space at x'. Adding them together is not mathematically defined independent of coordinates. We would need to have a way of mapping vectors from one tangent space to another. (We will later discuss one such mapping, called parallel transport, but it does not give a unique correspondence between tangent spaces.) Without such a map, the only tensor fields we can integrate on the manifold are scalar functions.

The second problem, which applies even to scalar functions, is that the integration measure d^4x is not coordinate-independent. Under the coordinate transformation $x^\alpha \to x'^\alpha(x)$, the volume element transforms with the determinant of the Jacobian matrix,

$$\mathrm{d}^4x \to \mathrm{d}^4x' = \det(M^\alpha{}_\beta)\mathrm{d}^4x \ . \tag{2.45}$$

We can cancel the change (2.45) by including the determinant of the metric in the volume element. Under a coordinate change the metric transforms as $g_{\alpha\beta} \to g'_{\alpha\beta} = (M^{-1})^\gamma{}_\alpha (M^{-1})^\delta{}_\beta g_{\gamma\delta}$, so its determinant transforms as

$$g \equiv \det(g_{\alpha\beta}) \to [\det(M^\alpha{}_\beta)]^{-2} \det(g_{\alpha\beta}) \ . \tag{2.46}$$

Noting that for a Lorentzian spacetime $g < 0$, the coordinate-invariant volume element (that reduces to the coordinate volume in the case of Minkowski space in Cartesian coordinates) is

$$\mathrm{d}V \equiv \mathrm{d}^4x\sqrt{-g} \ . \tag{2.47}$$

On a general manifold (where the metric may have Euclidean or Lorentzian signature), we write $\sqrt{|g|}$ instead of $\sqrt{-g}$. On a Lorentzian manifold with a diagonal metric, we have $\mathrm{d}V = \sqrt{-g_{00}}\mathrm{d}x^0\sqrt{g_{11}}\mathrm{d}x^1\sqrt{g_{22}}\mathrm{d}x^2\sqrt{g_{33}}\mathrm{d}x^3$. If we integrate over a line of constant coordinate, or an area or a three-volume defined similarly in terms of the coordinates, instead of a four-volume, we just pick the corresponding one, two or three coordinate differentials and metric components. For non-diagonal metrics, more sophisticated treatment is needed, but this will suffice for us.

The presence of the determinant is familiar from non-Cartesian coordinates in Euclidean space. For example, the metric of $\mathbb{R}^3$ in spherical coordinates is

$$\mathrm{d}s^2 = \mathrm{d}r^2 + r^2\mathrm{d}\theta^2 + r^2\sin^2\theta\mathrm{d}\varphi^2 \ , \tag{2.48}$$

so the integration measure is $r^2\sin\theta\mathrm{d}r\mathrm{d}\theta\mathrm{d}\varphi$.

2.4.5 Levi–Civita tensor

It is useful to introduce the **Levi–Civita tensor** on a general (four-dimensional, Lorentzian) manifold. It is defined as

$$\epsilon_{\alpha\beta\gamma\delta} \equiv \sqrt{-g}\eta_{\alpha\beta\gamma\delta} \ , \tag{2.49}$$

where $\eta_{\alpha\beta\gamma\delta}$ is the totally antisymmetric symbol defined by $\eta_{\alpha\beta\gamma\delta} = \eta_{[\alpha\beta\gamma\delta]}$ and $\eta_{0123} = 1$. The symbol $\eta_{\alpha\beta\gamma\delta}$ is not a tensor, and does not change under coordinate transformation. It has the same value whether the indices are up or down, just like the Kronecker delta $\delta_{\alpha\beta}$ and the symbol $\eta_{\alpha\beta}$.

2.4.6 Tensor components and observations

Tensor components depend on coordinates. What is their physical meaning, i.e. how do we relate them to observables? In the simplest case, observables are obtained from tensors by projecting with the appropriate physical vectors or tensors to get scalars (which are coordinate-independent). We will discuss this when we come to the **energy-momentum tensor** and the description of matter in GR in chapter 4. The relevant vectors include the time direction (four-velocity) of an observer and the spatial directions they consider, given by some locally orthonormal vectors. If the relevant physical quantities involve directions (as in the case of, say, forces), we will need to express them in the local coordinate system relevant to the observer we are considering. In a way, this is familiar from Newtonian mechanics. If we want to find how the acceleration vector $\vec{a}$ is related to what an observer measures, we project it with the basis vectors of the coordinate system the observer is using to find the acceleration in the relevant direction – or equivalently decompose the acceleration in terms of the relevant basis vectors. In Newtonian physics (in Cartesian coordinates, at least), this procedure is so simple that we don't have to think about it much.

In GR, things can be a little more complicated, but the idea is the same. We find a basis that is locally inertial at the point we want to consider (this is typically not a coordinate basis). The tensor components written in this basis represent physical quantities measured by the observer. If the metric is diagonal, the orthonormal basis vectors (denoted by hatted indices) $\underline{e}_{\hat{\alpha}}$ and $\tilde{e}^{\hat{\alpha}}$ are related to the coordinate basis vectors ∂_α and $\tilde{e}^\alpha$ by

$$\underline{e}_{\hat{\alpha}} = \frac{1}{\sqrt{|g_{\alpha\alpha}|}}\partial_\alpha \qquad \tilde{e}^{\hat{\alpha}} = \sqrt{|g_{\alpha\alpha}|}\tilde{e}^\alpha \qquad \text{(no sum over } \alpha) \ . \tag{2.50}$$

Given that

$$\underline{A} = A^\alpha \underline{e}_\alpha \qquad \tilde{\omega} = \omega_\alpha \tilde{e}^\alpha \qquad T = T^{\alpha\beta}\underline{e}_\alpha \underline{e}_\beta \ , \tag{2.51}$$

the components in the orthonormal frame are related to components in the original coordinate frame as

$$A^{\hat{\alpha}} = \sqrt{|g_{\alpha\alpha}|}A^\alpha \qquad \omega_{\hat{\alpha}} = \frac{1}{\sqrt{|g_{\alpha\alpha}|}}\omega_\alpha \qquad T^{\hat{\alpha}\hat{\beta}} = \sqrt{|g_{\alpha\alpha}|}\sqrt{|g_{\beta\beta}|}T^{\alpha\beta} \ . \tag{2.52}$$

If the metric is non-diagonal, more sophisticated machinery is needed; this simple treatment will be enough for us.

2.5 Back to the strong equivalence principle

2.5.1 How close to local Minkowski space?

Manifolds encode two important properties of GR: 1) there are no preferred coordinate systems, and 2) spacetime is locally Minkowski. We have discussed the first property, let us now turn to the second.

Consider the dot product of the basis vectors in a coordinate basis:

$$\underline{e}_\alpha \cdot \underline{e}_\beta = g_{\gamma\delta}(\underline{e}_\alpha)^\gamma(\underline{e}_\beta)^\delta = g_{\gamma\delta}\delta_\alpha{}^\gamma\delta_\beta{}^\delta = g_{\alpha\beta} \ . \tag{2.53}$$

We earlier noted that in general, the basis vectors in a coordinate basis are not orthonormal. The above relation quantifies this: the basis vectors in a coordinate basis are orthonormal everywhere in spacetime precisely when the spacetime is Minkowski and we use Cartesian coordinates.

Let us consider a bit more what it means for the spacetime to be globally Minkowski. The metric having the form $g_{\alpha\beta} = \eta_{\alpha\beta}$ is obviously a sufficient condition, but not a necessary one, as we can always write the metric in different coordinates. (Consider the merry-go-round coordinate system discussed in chapter 1.) In other words, the condition $g_{\alpha\beta} = \eta_{\alpha\beta}$ is coordinate-dependent. The necessary condition is that there exists a coordinate system where the metric is $g_{\alpha\beta} = \eta_{\alpha\beta}$. We want to have a coordinate-independent characterisation of this condition. In other words, we want to have a condition that is defined on the manifold and can thus be applied in any coordinate system.

As the first step, let us see what is the coordinate-dependence in the metric and its derivatives at a point. A general metric consists of 10 independent functions $g_{\alpha\beta}(x)$. A general coordinate transformation involves 4 functions $x'^\alpha(x)$ that we can choose at will. So the number of functions in the metric that are physically meaningful is 6. However we are here interested in the characterisation of the spacetime at a point, so the values of the functions and their derivatives are independent.

Choose an arbitrary point p. Consider an arbitrary coordinate transformation $x^\alpha \to x'^\alpha(x)$. Components of the metric transform with the inverse Jacobian matrix,

$$g_{\alpha\beta}(p) \to g'_{\alpha\beta}(p) = (M^{-1})^\mu{}_\alpha (M^{-1})^\nu{}_\beta g_{\mu\nu}|_p \equiv N^\mu{}_\alpha N^\nu{}_\beta g_{\mu\nu}|_p \ , \tag{2.54}$$

where we have introduced the notation $N^\alpha{}_\beta \equiv (M^{-1})^\alpha{}_\beta = \frac{\partial x^\alpha}{\partial x'^\beta}$ to make the equations easier to parse (as we will have quite a few inverse Jacobian matrices). Locally, at point p, the matrix $N^\alpha{}_\beta$ has 16 degrees of freedom (independent numbers): 4 for the coordinate values α, times 4 for the partial derivative directions β. Let us demand that $g'_{\alpha\beta}(p) = \eta_{\alpha\beta}$. Now $g_{\alpha\beta}(p)$ are 10 arbitrary numbers that we are given, and (2.54) equated to $\eta_{\alpha\beta}$ is a set of 10 quadratic equations for the 16 numbers $N^\alpha{}_\beta(p)$. A solution therefore exists, and we have 6 numbers left over. These parametrise precisely those transformations under which the metric $\eta_{\alpha\beta}$ transforms to itself, i.e. the Lorentz transformations. (As translations of the spacetime coordinates do not affect the Jacobian matrix, they can be added without changing the metric, to enlarge the 6-parameter group of Lorentz transformations to the 10-parameter group of Poincaré transformations.)

So we can always set the metric $g_{\alpha\beta}$ to $\eta_{\alpha\beta}$ at a point, i.e. make the basis vectors orthonormal there. What about the derivatives of the metric? The first derivative of the components of the metric (recall that it is not the components of a tensor) transforms as

$$\begin{aligned}\partial_\gamma g_{\alpha\beta}(p) \to \partial_\gamma g'_{\alpha\beta}(p) &= N^\rho{}_\gamma \partial_\rho (N^\mu{}_\alpha N^\nu{}_\beta g_{\mu\nu})|_p \\ &= N^\rho{}_\gamma \partial_\rho N^\mu{}_\alpha N^\nu{}_\beta g_{\mu\nu}|_p + N^\rho{}_\gamma N^\mu{}_\alpha \partial_\rho N^\nu{}_\beta g_{\mu\nu}|_p \\ &\quad + N^\rho{}_\gamma N^\mu{}_\alpha N^\nu{}_\beta \partial_\rho g_{\mu\nu}|_p \ . \end{aligned} \tag{2.55}$$

We now want to see whether we can set $\partial_\gamma g'_{\alpha\beta}(p) = 0$, given that we have already fixed $g_{\alpha\beta} = \eta_{\alpha\beta}$. So the numbers in both $g_{\alpha\beta}(p)$ and in $N^\alpha{}_\beta(p)$ are fixed[7], in addition to the new arbitrary numbers $\partial_\rho g_{\mu\nu}(p)$. We therefore have a set of linear equations for the numbers $\partial_\rho N^\nu{}_\beta(p)$. The number of equations is the number of independent components of $\partial_\rho g'_{\mu\nu}$. This is 40: 10 for the symmetric pair of numbers $\mu\nu$, times 4 for the directions of the partial derivative. The number of unknowns is the number of independent components of $\partial_\rho N^\nu{}_\beta = M^\alpha{}_\rho \frac{\partial^2 x^\nu}{\partial x'^\alpha \partial x'^\beta}$. As $M^\alpha{}_\rho$ is fixed by $N^\alpha{}_\rho$, the number of degrees of freedom is 40: 4 for the choice of coordinate direction ν, times 10 for the symmetric pair $\alpha\beta$ (partial derivatives commute). We have the same number of equations and unknowns, so we can set the first derivative of the metric to zero.

Coordinates where at point p

$$g_{\alpha\beta} = \eta_{\alpha\beta} \ , \qquad \partial_\gamma g_{\alpha\beta} = 0 \tag{2.56}$$

are called **locally inertial coordinates** at point p, and the associated basis vectors constitute a **local Lorentz frame**, also called a **local inertial frame**.

In the case of the metric at a point, we had 6 numbers left over, for its first derivative we were left with no extra parameters. We see that as we keep playing this game, we do worse every round, and will eventually lose. On the one hand, the number of equations we have to satisfy to make the nth derivative of the metric zero equals the number of degrees of freedom in $g_{\alpha\beta,\nu_1\ldots\nu_n}$. On the other hand, we have unknown numbers at our disposal equal to the number of degrees of freedom in $x^\alpha{}_{,\nu_1\ldots\nu_{n+1}}$ (where the derivatives are understood to be taken with respect to x'). Because the number of symmetric indices in the coordinate transformation is larger than in the derivatives of the metric, the number of free parameters in the former decreases with respect to those of the latter as n grows. At $n = 2$, we have $10 \times 10 = 100$ equations (for the two symmetric pairs of indices on $g_{\alpha\beta,\nu_1\nu_2}$), but only $4 \times 20 = 80$ unknowns (20 being the number of independent symmetric combinations $\nu_1\nu_2\nu_3$). So we can set 80 of the second derivatives of the metric to zero, but there remain 20 that we cannot eliminate.[8]

[7] Apart from the possibility to do Lorentz transformations, which however won't affect the argument.

[8] Our reasoning here applies only to a coordinate basis. It is possible to introduce a non-coordinate basis that is orthonormal at every point, $g_{\alpha\beta} = \eta_{\alpha\beta}$. Doing so means offloading the physics from the metric to the basis vectors. In a coordinate basis, the basis vectors are trivial and the metric is complicated, but we can instead make the metric trivial and the basis vectors complicated. We will not discuss this.

We can appreciate this result by analogy with setting up coordinates to describe a curve on $\mathbb{R}^2$. Consider a smooth curve on a flat plane. The curve exists independent of any coordinates, and we can choose Cartesian coordinates so that its description in terms of a function $f(x)$ around an arbitrary point p is simple. We can first set the zero of the y-axis to lie at p, so that $f|_p = 0$. This corresponds to choosing coordinates so that $g_{\alpha\beta}|_p = \eta_{\alpha\beta}$. We can then orient the coordinates so that the x-axis is tangent to the curve at p, so that $f'|_p = 0$. This corresponds to choosing coordinates so that $\partial_\gamma g_{\alpha\beta}|_p = 0$. The second derivative f'' then describes how strongly the curve opens up or down, or neither, at p. This curvature is an intrinsic property of the curve, and it cannot be put to zero by choice of coordinates.

Getting back to GR, we see that if the laws of non-gravitational physics (in practice, equations of motion) involve the spacetime only via the metric and its first derivative, they are locally the same as in SR, so the strong equivalence principle is satisfied: at a point, gravity has no effect. This is also true if they only depend on the 80 second derivatives that we can get rid of by a coordinate transformation. In contrast, if the equations of motion depend on any of the 20 second derivatives (or higher derivatives) that cannot be eliminated by a coordinate transformation, this is not the case, and gravity is locally measurable. This is the exact statement of the strong equivalence principle in GR. So calling it a principle is a bit misleading: it is a property that some non-gravitational laws of physics have and others don't have.

We started by defining manifolds as structures that are essentially local Minkowski spaces knitted together, and for which all coordinate descriptions are equally valid. We have now made this more precise: at a point the metric and its first derivative are those of Minkowski space (i.e. they have no physical degrees of freedom, only coordinate degrees of freedom), but this is true only for 80 out of the 100 degrees of freedom in the second derivative. So the second derivatives of the metric contain coordinate-independent information. It will turn out that the 20 second derivatives in fact contain all coordinate-independent local information about the spacetime in the sense that if and only if they are everywhere zero, the spacetime is Minkowski. In the next chapter we will find a coordinate-independent characterisation of these 20 degrees of freedom, that is, we will describe the **curvature** of the manifold in terms of tensors.