Differential and Riemannian Geometry for General Relativity: A Useful Toolbox

by an aspiring (but still amateur) geometrist

Denis Werth
International Center for Fundamental Physics, 2020-2021

Preface

We have been doing physics for several years now. Since then, physics was mostly Fourier transforms, solving differential equations and linear algebra. The question behind most of the physical problems was usually *How to diagonalize the Hamiltonian?* This is just because we did a lot of Quantum Physics. We were (and still are) used to the Dirac notations *i.e.* the bras and the kets, operators, commutators, etc. Even when it comes to contracting indices and using the metric tensor, Special Relativity gave us some ease. With General Relativity, the paradigm has changed, and so the mathematical formalism. Everything is continuous and formula are impossible to remember[1]. A lot of questions are raised. What is the tangent vector tangent to? What does the connection connect? Does a one-form act on a function? How to physically interpret the Riemann tensor? What is the link between the Lie derivative, Killing vectors and symmetries? Does a change of coordinates act at the same point? If you are just like me and all these questions prevent you from sleeping, these lecture notes are made for you.

What are these Lecture Notes? This handout presents the mathematics of General Relativity. It aims at offering the reader a mathematical toolbox and ease behind the concepts of Riemannian geometry. We will introduce the useful basic concepts of differential and Riemannian geometry, discuss their interpretations and give useful tips to easily apply these abstract concepts to physics. I will try to write these lecture notes so that you can have different reading levels. For example, you can entirely read the handout or just the essential at the end of each section. Several examples in violet also illustrate the abstract concepts throughout the script. We will not prove most of the results to keep these lecture notes relatively short. Finally, these lecture notes are mainly based on David Tong's lectures[2] and the heavy "Big Black Book"[3].

What These Lecture Notes are Not? Personally in General Relativity, I know how to apply formula, compute with several indices, integrate, change frames, etc. But I feel that I do not deeply understand the mathematical objects I am using. To help me understand the mathematics of General Relativity and to make things clear in my mind, I started to write these lecture notes. And I learned a lot! Obviously, the aim was not to write another lecture notes on General Relativity because many already exist on the web. Instead, I wanted to focus on mathematics.

First, I need to say that this is not a physics lecture. We will not discuss Einstein equations, present some solutions, linearize the theory to recover gravitational waves, etc... It means basically that we will not make a clear and deep link between geometry and gravity. But wait... Gravity is geometry! This rather simple statement gives me the opportunity to focus on mathematics and geometry instead of physics, hoping that physics will become more intuitive afterward. The rest is no longer my responsibility because (i) I am just a beginner *i.e.* I know pretty much nothing about General Relativity, and (ii) we have a wonderful teacher.

Finally, I would like to thank our teachers Marios Petropoulos and Philippe Grandclément for making me understand that I do not fully understand General Relativity.

[1] Can you write the torsion tensor in terms of components without looking at the formula?
[2] David Tong, *General Relativity*, University of Cambridge Part III Mathematical Tripos
[3] Charles W. Misner, Kip S. Thorne and John Archibald Wheeler, *Gravitation*

Contents

Chapter 1

Differential Geometry

The aim of this section is to fully understand the objects we manipulate in General Relativity. Most of the definitions are usually very formal and one loses intuition. Here, we give a meaningful sense of how different objects should be seen or at least imagined. Even if our discussion of differential geometry is not particularly rigorous, we will be careful about building up the mathematical objects in the right logical order.

1.1 Differential Manifolds

The first basic notion that we introduce is the concept of **manifold**. As most physicists are satisfied with a more fuzzy and intuitive definition, we will not discuss some elaborate framework. Instead, we should think of a manifold as a curved D-dimensional space so that if you zoom in to any patch, the manifold looks like $\mathbb{R}^D$. Viewed more globally, the manifold may have interesting curvature or topology.

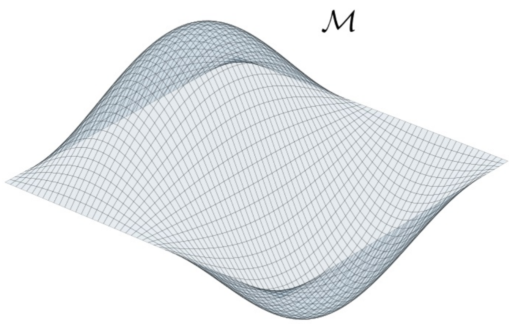


Figure 1.1. Example of a differential manifold in two dimensions which locally looks like $\mathbb{R}^2$.

If one wants to describe physics with manifolds, one needs to make manifolds differential. In simple words, a differential manifold $\mathcal{M}$ means that we can continuously (smoothly) go from one point $p \in \mathcal{M}$ of the manifold to another point $q \in \mathcal{M}$, and that $\mathcal{M}$ locally looks like $\mathbb{R}^D$. An illustration is provided above.

Example. Some simple examples in mathematics include Euclidean space $\mathbb{R}^D$, the sphere $\mathbb{S}^D$ and the torus $\mathbb{T}^D = \mathbb{S}^1 \times ... \times \mathbb{S}^1$. In statistical physics, the phase space of N particles is a $6N$-dimensional manifold.

The advantage of locally identify a differential manifold $\mathcal{M}$ to $\mathbb{R}^D$ is that we can now import our knowledge of how to do maths on $\mathbb{R}^D$. For example, we now how to parametrize a curve or to differentiate functions on $\mathbb{R}^D$.

A differential manifold is a smooth curved D-dimensional space that is locally Euclidean.

1.2 Vectors Redefined into Tangent Vectors

We know what vectors are from undergraduate physics, and we know what differential operators are. But we are not used to equating the two. Here, we give the tools to understand in what sense a vector should now be seen as a tangent vector *i.e.* a differential operator.

In classical mechanics working in a flat space, we describe the position of a particle as a vector $\boldsymbol{x} \in \mathbb{R}^3$ defined as an "arrow" drawn from some origin to a point. This notion does not generalize to other manifolds. For example, a line connecting two points on a sphere is not a vector. In general, there is no way to think of a point $p \in \mathcal{M}$ as a vector.

To first understand tangent vectors at a given point $p \in \mathcal{M}$, we would like to give a coordinate independent (formal) definition of differentiation, and then describe it using coordinates. Let us denote $\mathcal{C}^\infty(\mathcal{M})$ the set of all smooth functions over a manifold $\mathcal{M}$.

Definition. A **tangent vector** $\boldsymbol{X}_p$ is an object that differentiates functions at a point $p \in \mathcal{M}$. Specifically, $\boldsymbol{X}_p : \mathcal{C}^\infty(\mathcal{M}) \to \mathbb{R}$ satisfying

(i) $\boldsymbol{X}_p(f+g) = \boldsymbol{X}_p(f) + \boldsymbol{X}_p(g)$ for all $f, g \in \mathcal{C}^\infty(\mathcal{M})$.

(ii) $\boldsymbol{X}_p(f) = 0$ when f is a constant function.

(iii) $\boldsymbol{X}_p(fg) = f(p)\boldsymbol{X}_p(g) + \boldsymbol{X}_p(f)g(p)$ for all $f, g \in \mathcal{C}^\infty(\mathcal{M})$.

The last property is of course the Leibniz rule. Intuitively, tangent vectors tell us how things change in a given direction. They do this by differentiating. It is easy to check that the objects

$$\partial_\mu \Big|_p = \frac{\partial}{\partial x^\mu}\Big|_p, \tag{1.1}$$

which act on functions obeys all the requirements of a tangent vector[1]. However, they are much more than simple tangent vectors.

Indeed, the set of all tangent vectors at a point $p \in \mathcal{M}$ forms an D-dimensional vector space called the **tangent space** $\mathcal{T}_p(\mathcal{M})$. The tangent vectors $\partial_\mu\Big|_p$ provide a basis for $\mathcal{T}_p(\mathcal{M})$. This means that we can rewrite any tangent vector as

$$\boldsymbol{X}_p = X^\mu \partial_\mu\Big|_p, \tag{1.2}$$

with X^μ the components of the tangent vector in this basis. Note that to be fully rigorous regarding the notation, one needs to write $\boldsymbol{\partial}_\mu$ instead of ∂_μ because this object is a tangent vector.

What is it tangent to? So far, we have not really explained where the name "tangent vector" comes from. Consider a smooth curve in $\mathcal{M}$ that passes through the point $p \in \mathcal{M}$. We describe the curve by some coordinates $x^\mu(\lambda)$ where λ parameterizes the curve such that $\lambda = 0$ at $p \in \mathcal{M}$. Before we learned any differential geometry, we would say that the tangent vector to the curve at $\lambda = 0$ is

$$X^\mu = \frac{\mathrm{d}x^\mu(\lambda)}{\mathrm{d}\lambda}\Big|_{\lambda=0}. \tag{1.3}$$

But we can take these to be the components of the tangent vector $\boldsymbol{X}_p$ which we define as

$$\boldsymbol{X}_p = \frac{\mathrm{d}x^\mu(\lambda)}{\mathrm{d}\lambda}\Big|_{\lambda=0} \partial_\mu\Big|_{\lambda=0}. \tag{1.4}$$

The tangent vector tells us how fast any function $f \in \mathcal{C}^\infty(\mathcal{M})$ changes as we move along the curve. This gives also meaning to the term "tangent space" for $\mathcal{T}_p(\mathcal{M})$. It is literally the space of all possible tangents to curves passing through the point $p \in \mathcal{M}$. An illustration of a two dimensional manifold embedded in $\mathbb{R}^3$ is shown below. Note that the tangent spaces $\mathcal{T}_p(\mathcal{M})$ and $\mathcal{T}_q(\mathcal{M})$ at different points $p \neq q$ are different. There is no way to compare vectors in $\mathcal{T}_p(\mathcal{M})$ and vectors in $\mathcal{T}_q(\mathcal{M})$.

So far we have only defined tangent vectors at a point p. It is useful to consider objects in which there is a choice of tangent vector for every point $p \in \mathcal{M}$. We call objects that vary over space fields. A **vector field** $\boldsymbol{X}$ is defined to be a smooth assignment of a tangent vector $\boldsymbol{X}_p$ to each point $p \in \mathcal{M}$. Given a coordinate basis, we can expand any vector field as

$$\boldsymbol{X} = X^\mu \partial_\mu, \tag{1.5}$$

where the X^μ are now smooth functions on $\mathcal{M}$. This tangent vector field $\boldsymbol{X}$, given a certain curve parameterized by $x^\mu(\lambda)$, acts on a function as a differential operator

$$\boldsymbol{X}(f) = X^\mu \partial_\mu f = \frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda}\frac{\partial f}{\partial x^\mu} = \frac{\mathrm{d}f(\lambda)}{\mathrm{d}\lambda}. \tag{1.6}$$

[1] Note that the index μ is a subscript rather than superscript that we use for the coordinates x^μ.

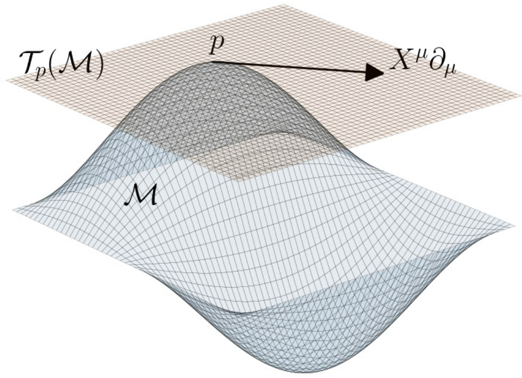


Figure 1.2. Illustration of a tangent vector space $\mathcal{T}_p(\mathcal{M})$ at the point $p \in \mathcal{M}$ with a tangent vector $\boldsymbol{X} = X^\mu\partial_\mu$ in a two dimensional manifold.

It means that a tangent vector acting on a function tells us about how this function vary along the curve.

At any point $p \in \mathcal{M}$ on the manifold, we define a tangent vector $\boldsymbol{X}_p : \mathcal{C}^\infty(\mathcal{M}) \to \mathbb{R}$ that satisfies properties of differentiation (linearity, Leibniz rule, ...). The set of objects $\partial_\mu\Big|_p$ forms a basis of the tangent space $\mathcal{T}_p(\mathcal{M})$. Thus the tangent vector can be decomposed as

$$\boldsymbol{X}_p = X^\mu\partial_\mu\Big|_p. \tag{1.7}$$

Allowing the components X^μ to vary on $\mathcal{M}$, we define a vector field $\boldsymbol{X} : \mathcal{C}^\infty(\mathcal{M}) \to \mathcal{C}^\infty(\mathcal{M})$ as

$$\boldsymbol{X} = X^\mu\partial_\mu. \tag{1.8}$$

The action of a tangent vector field $\boldsymbol{X}$, given a certain curve parameterized by $x^\mu(\lambda)$, on a function f is

$$\boldsymbol{X}(f) = X^\mu\partial_\mu f = \frac{\mathrm{d}f(\lambda)}{\mathrm{d}\lambda}, \tag{1.9}$$

which is a number when evaluated at a point $p \in \mathcal{M}$.

1.3 Integral Curves

The last section was formal but essential to understand that a tangent vector is a differential object. Fortunately, there is a slightly different way of thinking about vector fields on a manifold. This point of view is the one I use when I think of tangent vectors.

We have seen previously that the components of a tangent vector are

$$X^\mu = \frac{\mathrm{d}x^\mu(\lambda)}{\mathrm{d}\lambda}. \tag{1.10}$$

It means that to compute a tangent vector $\boldsymbol{X}_p$ at a given point $p \in \mathcal{M}$, one must specify a curve passing through p. As the curve is defined by $x^\mu(\lambda)$, one must differentiate each coordinate with respect to λ to recover a tangent vector.

Example. Let us consider the curve $\mathcal{C} : \{\theta = \pi/2, \varphi = \lambda\}$ on the two dimensional sphere $\mathbb{S}^2$ that we parameterize with θ and φ. This curve clearly is the equator. To explicitly compute a tangent vector at some point, one needs to specify a basis. We choose $\{\partial_\theta, \partial_\varphi\}$. Then our vector $\boldsymbol{X}$ is written

$$\boldsymbol{X} = \begin{pmatrix} \frac{\mathrm{d}\theta}{\mathrm{d}\lambda} \\ \frac{\mathrm{d}\varphi}{\mathrm{d}\lambda} \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \partial_\varphi. \tag{1.11}$$

The previous example shows that one can explicitly write a tangent vector if a specific curve is given. Indeed, a tangent vector at a given point can be oriented in different ways.

Alternatively, given a vector field $\boldsymbol{X} = X^\mu \partial_\mu$, we can integrate the differential equation (1.10) subject to some initial conditions. The generated streamlines are called **integral curves**. The integral lines are then simply describe by

$$x^\mu(\lambda) = x^\mu(0) + \lambda X^\mu + \mathcal{O}(\lambda^2), \tag{1.12}$$

that illustrates how the tangent vector components give the variation along a curve.

Example. • Consider again the sphere $\mathbb{S}^2$ in polar coordinates with a vector $\boldsymbol{X} = \partial_\varphi$. The integral curves solve the equation (1.10), which are

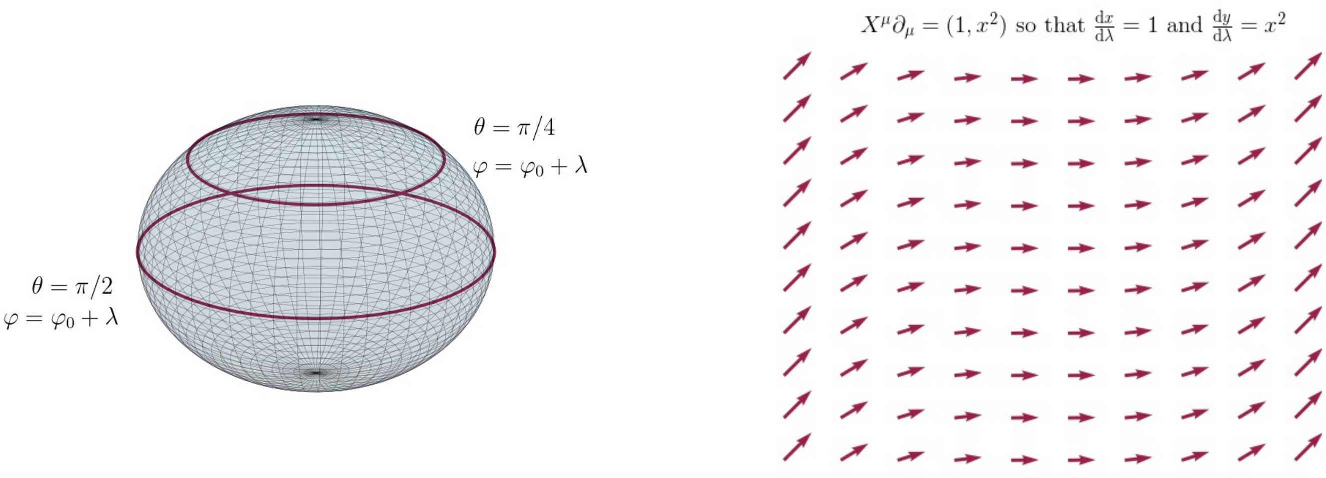
$$\frac{\mathrm{d}\theta}{\mathrm{d}\lambda} = 0 \quad \text{and} \quad \frac{\mathrm{d}\varphi}{\mathrm{d}\lambda} = 1, \tag{1.13}$$

that has the solution $\theta = \theta_0$ and $\varphi = \varphi_0 + \lambda$. The integral lines are shown below.

• Consider the vector field on $\mathbb{R}^2$ with Cartesian coordinates $\boldsymbol{X} = (1, x^2) = \partial_x + x^2\partial_y$. The equation for the integral curves is now

$$\frac{\mathrm{d}x}{\mathrm{d}\lambda} = 1 \quad \text{and} \quad \frac{\mathrm{d}y}{\mathrm{d}\lambda} = x^2, \tag{1.14}$$

that has the solution $x(\lambda) = x_0 + \lambda$ and $y(\lambda) = y_0 + \frac{1}{3}(x_0 + \lambda)^3$. The associated flow lines are shown below.



(a) $\boldsymbol{X} = X^\mu \partial_\mu = \partial_\varphi$.

(b) $\boldsymbol{X} = X^\mu \partial_\mu = \partial_x + x^2 \partial_y$.

Figure 1.3. Some integral curves corresponding to the two examples given above.

Given a curve $\mathcal{C}$ parameterized by $x^\mu(\lambda)$, one can compute the components of a tangent vector X^μ using differentiation[a]

$$X^\mu = \frac{\mathrm{d}x^\mu(\lambda)}{\mathrm{d}\lambda}. \tag{1.15}$$

Alternatively, given a tangent vector $\boldsymbol{X} = X^\mu \partial_\mu$, one can integrate (1.15) and obtain integral curves[b] parameterized by $x^\mu(\lambda)$. A tangent vector tells us about variation along a curve, as shown by

$$x^\mu(\lambda) = x^\mu(0) + \lambda X^\mu + \mathcal{O}(\lambda^2). \tag{1.16}$$

A tangent vector must be seen as a directional derivative operator along a curve.

[a] Note that this is possible if one first chooses a basis and a set of coordinates.
[b] Note that in practice, only very simple tangent vectors can be integrated.

1.4 Transformation Law for Tangent Vectors

An especially useful basis in the tangent space at a point[2] $p \in \mathcal{M}$ is induced by any coordinate system

$$\boldsymbol{e}_\mu = \frac{\partial}{\partial x^\mu} = \partial_\mu, \tag{1.17}$$

that must be seen as a directional derivative along a curve. A transformation from one basis to another in the same tangent space $\mathcal{T}_p(\mathcal{M})$ at the event $p \in \mathcal{M}$ is produced by a matrix

[2] It is more appropriate to call a point an **event**.

$$e'_\nu = \frac{\partial}{\partial x'^\nu} = \frac{\partial}{\partial x^\mu}\frac{\partial x^\mu}{\partial x'^\nu} = e_\mu J^\mu_\nu. \tag{1.18}$$

The matrix $\boldsymbol{J}$ is the **Jacobian** of the transformation $x^\mu \to x'^\mu$. One must see this transformation as the **chain rule**. The components of a tangent vector must transform by the inverse matrix

$$X'^\nu = (J^{-1})^\nu_\mu X^\mu. \tag{1.19}$$

Note that we have $(J^{-1})^\nu_\mu J^\mu_\tau = \delta^\nu_\tau$. This inverse transformation law guarantees compatibility between the expressions $\boldsymbol{X} = X'^\nu e'_\nu$ and $\boldsymbol{X} = X^\mu e_\mu$

$$\boldsymbol{X} = X'^\nu e'_\nu = ((J^{-1})^\nu_\tau X^\tau)(e_\mu J^\mu_\nu) = X^\tau((J^{-1})^\nu_\tau J^\mu_\nu) e_\mu = X^\tau \delta^\mu_\tau e_\mu = X^\mu e_\mu. \tag{1.20}$$

Example. Let us consider the two dimensional flat plane[3]. One can use polar coordinates $x^\mu : \{r, \theta\}$ or Cartesian coordinates $x'^\mu : \{x = r\cos\theta, y = r\sin\theta\}$ to define an event, a curve, etc. The Jacobian of the $x^\mu \to x'^\mu$ transformation is defined as

$$\boldsymbol{J} = \begin{pmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{pmatrix} \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}, \qquad \boldsymbol{J}^{-1} = \frac{1}{r}\begin{pmatrix} r\cos\theta & r\sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}. \tag{1.21}$$

Now if we consider a tangent vector $\boldsymbol{X} = (0, 1) = \partial_\theta$ in polar coordinates, one can rewrite[4] it in Cartesian coordinates

$$\partial_\theta = J^\mu_\theta \partial_\mu = J^x_\theta \partial_x + J^y_\theta \partial_y = -r\sin\theta\partial_x + r\cos\theta\partial_y = -y\partial_x + x\partial_y. \tag{1.22}$$

Our tangent vector $\boldsymbol{X}$ is written $\boldsymbol{X} = (-y, x)$ in Cartesian coordinates. Notice that in Cartesian coordinates, it is not obvious that our tangent vector gives rise to integral curves rotating around the origin.

The change of basis and coordinates $x^\mu \to x'^\mu$ of a tangent vector $\boldsymbol{X}$ is made by the use of the Jacobian[a][b]

$$J^\mu_\nu = \frac{\partial x^\mu}{\partial x'^\nu}, \qquad (J^{-1})^\mu_\nu = \frac{\partial x'^\mu}{\partial x^\nu} \quad \text{satisfying} \quad J^\nu_\mu J^\mu_\tau = \delta^\nu_\tau. \tag{1.23}$$

The basis tangent vectors $e_\mu = \partial_\mu$ transform with $\boldsymbol{J}$

$$e'_\nu = e_\mu J^\mu_\nu. \tag{1.24}$$

The tangent vector coordinates X^μ, depending on the choice of coordinates, transform with $\boldsymbol{J}^{-1}$

$$X'^\mu = (J^{-1})^\mu_\nu X^\nu. \tag{1.25}$$

[a] Note that the equations give the components of $\boldsymbol{J}$ and $\boldsymbol{J}^{-1}$.

[b] We remember this transformation as differentiating the new coordinates with respect to the old coordinates.

[3] We eliminate the origin because it is a singular point for polar coordinates.

[4] Note that because we transform a basis tangent vector, we use the Jacobian and not its inverse.

1.5 One-Forms

We have seen that tangent vectors at an event $p \in \mathcal{M}$ live in the tangent space $\mathcal{T}_p(\mathcal{M})$. Since $\mathcal{T}_p(\mathcal{M})$ is a vector space (one can choose $\{\partial_\mu\}$ as a basis), there exists a dual vector space to $\mathcal{T}_p(\mathcal{M})$ denoted $\mathcal{T}_p^\star(\mathcal{M})$. This space is called the **cotangent space** and its elements are linear functions from $\mathcal{T}_p(\mathcal{M})$ to $\mathbb{R}$.

An element $\boldsymbol{\omega} \in \mathcal{T}_p^\star(\mathcal{M})$ of the cotangent space is called a **one-form**. A one-form acts on a tangent vector to give a number. This procedure must be seen as a scalar product[5]

$$\begin{aligned} \boldsymbol{\omega} \in \mathcal{T}_p^\star(\mathcal{M}) : \mathcal{T}_p(\mathcal{M}) &\to \mathbb{R} \\ \boldsymbol{X} &\to \langle \boldsymbol{\omega} | \boldsymbol{X} \rangle \end{aligned} \tag{1.26}$$

The simplest example of a one-form is the differential $\mathrm{d}f$ of a function f. The action of a tangent vector $\boldsymbol{X}$ on f being $\boldsymbol{X}(f) = X^\mu \partial_\mu f$, the action of $\mathrm{d}f$ on $\boldsymbol{X}$ is

$$\langle \mathrm{d}f | \boldsymbol{X} \rangle = \boldsymbol{X}(f) = X^\mu \partial_\mu f. \tag{1.27}$$

Noting that $\mathrm{d}f$ is expressed in terms of the coordinates as $\mathrm{d}f = \partial_\mu f \mathrm{d}x^\mu$, it is natural to regard $\{\mathrm{d}x^\mu\}$ as a basis for $\mathcal{T}_p^\star(\mathcal{M})$. Moreover, we have

$$\langle \mathrm{d}x^\mu | \partial_\nu \rangle = \frac{\partial x^\mu}{\partial x^\nu} = \delta^\mu_\nu. \tag{1.28}$$

An arbitrary one-form $\boldsymbol{\omega}$ can then be written

$$\boldsymbol{\omega} = \omega_\mu \mathrm{d}x^\mu, \tag{1.29}$$

where ω_μ are the components of $\boldsymbol{\omega}$. Given a tangent vector $\boldsymbol{X}$, one can act on this tangent vector with the one-form[6]

$$\langle \boldsymbol{\omega} | \boldsymbol{X} \rangle = \langle \omega_\mu \mathrm{d}x^\mu | X^\nu \partial_\mu \rangle = \omega_\mu X^\nu \langle \mathrm{d}x^\mu | \partial_\mu \rangle = \omega_\mu X^\nu \delta^\mu_\nu = \omega_\mu X^\mu. \tag{1.30}$$

From $\mathrm{d}x'^\nu = \frac{\partial x'^\nu}{\partial x^\mu} \mathrm{d}x^\mu$, we find the coordinate transformation of the one-form components

$$\omega'_\nu = \frac{\partial x'^\mu}{\partial x^\nu} \omega_\mu \tag{1.31}$$

[5] We draw an analogy with quantum mechanics when we write $\langle \phi | \psi \rangle$ with $|\psi\rangle \in \mathcal{H}$ and $\langle \phi | \in \mathcal{H}^\star$, the Hilbert space and its dual space respectively.

[6] Note that the scalar product is defined between a tangent vector and a one-form and not between two tangent vectors or two one-forms.

A one-form $\boldsymbol{\omega}$ is an element of the dual vector space $\mathcal{T}_p^\star(\mathcal{M})$ that acts on a tangent vector $\boldsymbol{X}$ to give a number[a] $\langle \boldsymbol{\omega} | \boldsymbol{X} \rangle = \omega_\mu X^\mu$. Any one-form $\boldsymbol{\omega}$ can be decomposed[b] on the basis $\{\mathrm{d}x^\mu\}$ as

$$\boldsymbol{\omega} = \omega_\mu \mathrm{d}x^\mu, \tag{1.32}$$

where ω_μ are the components of $\boldsymbol{\omega}$. The basis elements satisfy

$$\langle \mathrm{d}x^\mu | \partial_\nu \rangle = \frac{\partial x^\mu}{\partial x^\nu} = \delta^\mu_\nu. \tag{1.33}$$

[a]This action defines an inner product.

[b]Remember that one-form components have low indices whereas tangent vector components have upper indices.

1.6 Tensors

A **tensor** of type (q, r) is a multilinear object which maps q elements of $\mathcal{T}_p^\star(\mathcal{M})$ and r elements of $\mathcal{T}_p(\mathcal{M})$ to a number. Such an object, called a tensor of rank $q + r$, is written in terms of the bases described earlier as

$$\boldsymbol{T} = T^{\mu_1 \dots \mu_q}{}_{\nu_1 \dots \nu_r}\, \partial_{\mu_1} \dots \partial_{\mu_q}\, \mathrm{d}x^{\nu_1} \dots \mathrm{d}x^{\nu_r}. \tag{1.34}$$

Note that we deliberately write the string of lower indices after the upper indices, and the tensor product between basis elements is not written. In some sense this is unnecessary, and we do not lose any information by writing $T^{\mu_1 \dots \mu_q}_{\nu_1 \dots \nu_r}$. Nonetheless, we will see later that it is a useful habit to get into.

Example. A tangent vector is a tensor of type[7] $(1, 0)$ and a one-form is a tensor of type $(0, 1)$.

Let $\boldsymbol{X}_i = X_i^\mu \partial_\mu$ with $1 \leq i \leq r$ and $\boldsymbol{\omega}_j = \omega_{j\,\mu} \mathrm{d}x^\mu$ with $1 \leq j \leq q$ be i tangent vectors and j one-forms. The action of a tensor $\boldsymbol{T}$ on them yields a number

$$\boldsymbol{T}(\boldsymbol{\omega}_1, \dots, \boldsymbol{\omega}_q; \boldsymbol{X}_1, \dots, \boldsymbol{X}_r) = T^{\mu_1 \dots \mu_q}{}_{\nu_1 \dots \nu_r}\, \omega_{\mu_1} \dots \omega_{\mu_q}\, X^{\nu_1} \dots X^{\nu_r}. \tag{1.35}$$

As with vector fields and one-forms, we can ask how the components of a tensor transform. We know that tangent vector basis elements ∂_μ and one-form basis elements $\mathrm{d}x^\mu$ transform as

$$\partial'_\nu = J^\mu_\nu \partial_\mu \quad \text{and} \quad \mathrm{d}x'^\nu = (J^{-1})^\nu_\mu \mathrm{d}x^\mu. \tag{1.36}$$

Then, it is clear that the lower components of a tensor transform by multiplying by $\boldsymbol{J}$ and the upper components by multiplying by $\boldsymbol{J}^{-1}$. So, for example, a rank $(1, 2)$ tensor transforms as

$$T'^\mu{}_{\rho\nu} = (J^{-1})^\mu_\sigma\, J^\tau_\rho\, J^\lambda_\nu\, T^\sigma{}_{\tau\lambda} = \frac{\partial x^\mu}{\partial x'^\sigma} \frac{\partial x'^\tau}{\partial x^\rho} \frac{\partial x'^\lambda}{\partial x^\nu} T^\sigma{}_{\tau\lambda}, \tag{1.37}$$

[7]Using the fact that $\mathcal{T}_p^{\star\star}(\mathcal{M}) = \mathcal{T}_p(\mathcal{M})$.

for a $x^\mu \to x'^\mu$ coordinate transformation. Note that we usually do not write the prime because there is no ambiguity regarding the change of coordinates. This transformation is just the chain rule and acts at two different points.

Similarly to tangent vectors fields, we can define tensor fields by making each components of the tensor a function that can be smoothly evaluated on the manifold. Note that on a manifold of dimension D, a tensor $\boldsymbol{T}$ of type (q, r) has D^{q+r} components. For a tensor field, each of these is a function over $\mathcal{M}$.

In General Relativity, every equation satisfies the **principle of general covariance** *i.e.* the equation preserves its form under a general coordinate transformation. As one can define a tensor independently of the basis that we chose, a tensorial equation is valid in any coordinate system *i.e.* in any frame. Note that, in particular, a tensor is zero (at a point) in one coordinate system if and only if the tensor is zero (at the same point) in another coordinate system.

A tensor $\boldsymbol{T}$ of type (q, r) acts on q one-formes and r tangent vectors to give a number

$$\boldsymbol{T}(\boldsymbol{\omega}_1, ..., \boldsymbol{\omega}_q; \boldsymbol{X}_1, ..., \boldsymbol{X}_r) = T^{\mu_1 \dots \mu_q}{}_{\nu_1 \dots \nu_r}\, \omega_{\mu_1} \dots \omega_{\mu_q} X^{\nu_1} \dots X^{\nu_r}. \tag{1.38}$$

Under a coordinate transformation, the lower components of a tensor transform by multiplying by $\boldsymbol{J}$ and the upper components by multiplying by $\boldsymbol{J}^{-1}$. Intuitively, this is the chain rule. For example,

$$T'^{\mu}{}_{\rho\nu} = \frac{\partial x^\mu}{\partial x'^\sigma} \frac{\partial x'^\tau}{\partial x^\rho} \frac{\partial x'^\lambda}{\partial x^\nu} T^\sigma{}_{\tau\lambda}. \tag{1.39}$$

The principle of general covariance tells us that every physical equation should be written in a tensorial form such that it is valid in every coordinate system.

1.7 Operations on Tensor Fields

There are a number of operations that we can do on tensor fields. The basis algebraic operations are the following.

Linear combination. Given two (q, r) tensors $\boldsymbol{A}, \boldsymbol{B}$ and two scalars α, β, their linear combination $\boldsymbol{T} = \alpha \boldsymbol{A} + \beta \boldsymbol{B}$ is also a (q, r) tensor. In terms of components, this reads

$$T^{\mu_1 \dots \mu_q}{}_{\nu_1 \dots \nu_r} = \alpha A^{\mu_1 \dots \mu_q}{}_{\nu_1 \dots \nu_r} + \beta B^{\mu_1 \dots \mu_q}{}_{\nu_1 \dots \nu_r}. \tag{1.40}$$

Tensor product. Given a (q, r) tensor $\boldsymbol{A}$ and a (q', r') tensor $\boldsymbol{B}$, their tensor product $\boldsymbol{T} = \boldsymbol{A} \otimes \boldsymbol{B}$ is a $(q + q', r + r')$ tensor. In terms of components, it reads

$$T^{\mu_1 \dots \mu_{q+q'}}{}_{\nu_1 \dots \nu_{r+r'}} = A^{\mu_1 \dots \mu_q}{}_{\nu_1 \dots \nu_r} B^{\mu_{q+1} \dots \mu_{q+q'}}{}_{\nu_{r+1} \dots \nu_{r+r'}}. \tag{1.41}$$

Contraction. Given a (q, r) tensor, one can associate to it a $(q - 1, r - 1)$ tensor via contraction of one upper and one lower index

$$T^{\mu_1\dots\mu_q}{}_{\nu_1\dots\nu_r} \to T^{\mu_1\dots\mu_{q-1}}{}_{\nu_1\dots\nu_{r-1}} = T^{\mu_1\dots\mu_{q-1}\lambda}{}_{\nu_1\dots\nu_{r-1}\lambda}. \tag{1.42}$$

Note that contraction over different pairs of indices will in general give rise to different tensors. For example, $T^{\mu}_{\nu\mu}$ and $T^{\mu}_{\mu\nu}$ will be different in general.

Symmetrisation and anti-Symmetrisation. Given a $(0,2)$ tensor $\boldsymbol{T} = T_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu$, one can decompose it into its symmetric and anti-symmetric parts as

$$\boldsymbol{T}(\boldsymbol{X},\boldsymbol{Y}) = \boldsymbol{S}(\boldsymbol{X},\boldsymbol{Y}) + \boldsymbol{A}(\boldsymbol{X},\boldsymbol{Y}) \quad \text{with} \quad \begin{aligned} \boldsymbol{S}(\boldsymbol{X},\boldsymbol{Y}) &= \frac{1}{2}\left[\boldsymbol{T}(\boldsymbol{X},\boldsymbol{Y}) + \boldsymbol{T}(\boldsymbol{Y},\boldsymbol{X})\right] \\ \boldsymbol{A}(\boldsymbol{X},\boldsymbol{Y}) &= \frac{1}{2}\left[\boldsymbol{T}(\boldsymbol{X},\boldsymbol{Y}) - \boldsymbol{T}(\boldsymbol{Y},\boldsymbol{X})\right] \end{aligned} \tag{1.43}$$

In index notation, this becomes

$$S_{\mu\nu} = \frac{1}{2}(T_{\mu\nu} + T_{\nu\mu}) \quad \text{and} \quad A_{\mu\nu} = \frac{1}{2}(T_{\mu\nu} - T_{\nu\mu}), \tag{1.44}$$

which is just like taking the symmetric and anti-symmetric part of a matrix. These operations being frequently used, we introduce some new notation. We define

$$T_{(\mu\nu)} = \frac{1}{2}(T_{\mu\nu} + T_{\nu\mu}) \quad \text{and} \quad T_{[\mu\nu]} = \frac{1}{2}(T_{\mu\nu} - T_{\nu\mu}). \tag{1.45}$$

Note that a product of a symmetric and an anti-symmetric tensor is always zero. Indeed, let us consider $\boldsymbol{S}$ a symmetric tensor ($S_{\mu\nu} = S_{\nu\mu}$) and $\boldsymbol{A}$ an anti-symmetric tensor ($A_{\mu\nu} = -A_{\nu\mu}$). Then,

$$S^{\mu\nu}A_{\mu\nu} = -S^{\nu\mu}A_{\nu\mu} = -S^{\mu\nu}A_{\mu\nu}, \tag{1.46}$$

where in the last step, we rewrite the indices because these are dummy indices. We finally obtain $S^{\mu\nu}A_{\mu\nu} = 0$.

Example. • These operations generalise to other tensors. For a $(3,1)$ tensor symmetric under the two first upper indices, one obtains

$$T^{(\mu\nu)\rho}{}_{\sigma} = \frac{1}{2}(T^{\mu\nu\rho}{}_{\sigma} + T^{\nu\mu\rho}{}_{\sigma}). \tag{1.47}$$

• Similarly, for a totally symmetric $(0,3)$ tensor, one obtains

$$T_{(\mu\nu\lambda)} = \frac{1}{3!}(T_{\mu\nu\lambda} + T_{\mu\lambda\nu} + T_{\nu\mu\lambda} + T_{\nu\lambda\mu} + T_{\lambda\mu\nu} + T_{\lambda\nu\mu}), \tag{1.48}$$

where we considered all permutations of three elements (hence the 3!). For a totally anti-symmetric tensor, one obtains the same but with a minus sign when the permutation cannot be obtain from the ordered indices with a cyclic permutation

$$T_{[\mu\nu\lambda]} = \frac{1}{3!}(T_{\mu\nu\lambda} - T_{\mu\lambda\nu} - T_{\nu\mu\lambda} + T_{\nu\lambda\mu} + T_{\lambda\mu\nu} - T_{\lambda\nu\mu}). \tag{1.49}$$

• If a $(0,2)$ tensor $\boldsymbol{T}$ is totally anti-symmetric, then $T_{\mu\nu} = -T_{\nu\mu}$.

• The product of a totally symmetric tensor with a totally anti-symmetric tensor is zero. Indeed, let us consider $A^{\mu\nu}$ symmetric and $B_{\mu\nu}$ anti-symmetric. Then one obtains

$$A^{\mu\nu}B_{\mu\nu} = -A^{\nu\mu}B_{\nu\mu} = -A^{\mu\nu}B_{\mu\nu}, \tag{1.50}$$

because μ and ν are dummy variables.

How to prove that an object is a tensor? To prove that an object is a tensor, one has to verify that the object (its components) transforms as a tensor. Schematically, if one obtains

$$\boldsymbol{T}' = (...) \, \boldsymbol{T} + \text{junk}, \tag{1.51}$$

then the object $\boldsymbol{T}$ is not a tensor. Note that when one wants to manipulate tensors, it is usually easier to deal with its components.

A linear combination of tensors is a same type tensor.

A tensor product between a (q, r) tensor and a (q', r') tensor gives a $(q + q', r + r')$ tensor.

Given a (q, r) tensor, one can associate to it a $(q - 1, r - 1)$ tensor via contraction of one upper and one lower index[a].

Given a tensor, one can define the symmetric and the anti-symmetric part of the tensor

$$T_{(\mu\nu)} = \frac{1}{2}(T_{\mu\nu} + T_{\nu\mu}) \quad \text{and} \quad T_{[\mu\nu]} = \frac{1}{2}(T_{\mu\nu} - T_{\nu\mu}). \tag{1.52}$$

These definitions are generalized to other tensors of higher rank. A product of a symmetric tensor with an anti-symmetric tensor is zero.

An object $\boldsymbol{T}$ that transforms as

$$\boldsymbol{T}' = (...) \, \boldsymbol{T} + \text{junk}, \tag{1.53}$$

is usually not a tensor.

[a] Note that $T^{\mu}_{\nu\mu}$ and $T^{\mu}_{\mu\nu}$ will be different in general.

1.8 Commutators and Lie Derivative

Given two tangent vector fields $\boldsymbol{X}$ and $\boldsymbol{Y}$, we cannot multiply them together to get a new tangent vector field. Roughly speaking, this is because the product $\boldsymbol{XY}$ is a second order differential operator rather a first order operator. This reveals itself in a failure of Leibnizarity for the object $\boldsymbol{XY}$,

$$\boldsymbol{XY}(fg) = \boldsymbol{X}(f\boldsymbol{Y}(g) + g\boldsymbol{Y}(f)) = \boldsymbol{X}(f)\boldsymbol{Y}(g) + f\boldsymbol{X}(\boldsymbol{Y}(g)) + g\boldsymbol{X}(\boldsymbol{Y}(f)) + \boldsymbol{X}(g)\boldsymbol{Y}(f), \tag{1.54}$$

which is not the same as $f\boldsymbol{XY}(g) + g\boldsymbol{XY}(f)$.

However, one can build a new tangent vector field by taking the **commutator** $[\boldsymbol{X},\boldsymbol{Y}]$, which acts on function f as

$$[\boldsymbol{X},\boldsymbol{Y}](f)=\boldsymbol{X}(\boldsymbol{Y}(f))-\boldsymbol{Y}(\boldsymbol{X}(f)). \tag{1.55}$$

This is also known as the **Lie bracket**. Evaluated in a coordinate basis, the commutator is given by

$$[\boldsymbol{X},\boldsymbol{Y}](f)=\left(X^\mu\frac{\partial Y^\nu}{\partial x^\mu}-Y^\mu\frac{\partial X^\nu}{\partial x^\mu}\right)\frac{\partial f}{\partial x^\nu}. \tag{1.56}$$

It is not difficult to check that the commutator obeys the **Jacobi identity**

$$[\boldsymbol{X},[\boldsymbol{Y},\boldsymbol{Z}]]+[\boldsymbol{Y},[\boldsymbol{Z},\boldsymbol{X}]]+[\boldsymbol{Z},[\boldsymbol{X},\boldsymbol{Y}]]=0, \tag{1.57}$$

which one can remember as the sum of the cyclic permutations of $\boldsymbol{X}$, $\boldsymbol{Y}$ and $\boldsymbol{Z}$ being zero. This ensures that the set of all tangent vector fields on a manifold $\mathcal{M}$ has the mathematical structure of a Lie algebra.

So far, we know how to differentiate a function f. This requires us to introduce a tangent vector field $\boldsymbol{X}$, and the new function $\boldsymbol{X}(f)$ can be viewed as the derivative of f in the direction of $\boldsymbol{X}$. But is it possible to differentiate a tangent vector field? Formally speaking, we need to subtract two tangent vector living on two different tangent spaces. This procedure is done using the **Lie derivative** denoted $\mathcal{L}_{\boldsymbol{X}}$. Without any construction, we give the definition of the Lie derivative.

The Lie derivative of a function f along $\boldsymbol{X}$ is

$$\mathcal{L}_{\boldsymbol{X}}f=\boldsymbol{X}(f)=X^\mu\frac{\partial f}{\partial x^\mu}. \tag{1.58}$$

In other words, acting on functions with the Lie derivative coincides with the action of the tangent vector field.

The Lie derivative of a tangent vector field $\boldsymbol{Y}$ along $\boldsymbol{X}$ is the commutator

$$\mathcal{L}_{\boldsymbol{X}}\boldsymbol{Y}=[\boldsymbol{X},\boldsymbol{Y}]. \tag{1.59}$$

Note the analogy with quantum mechanics and for example the angular momentum. Written in terms of components, it is

$$(\mathcal{L}_{\boldsymbol{X}}\boldsymbol{Y})^\mu=\mathcal{L}_XY^\mu=X^\nu\frac{\partial Y^\mu}{\partial x^\nu}-Y^\nu\frac{\partial X^\mu}{\partial x^\nu}=X^\nu\partial_\nu Y^\mu-Y^\nu\partial_\nu X^\mu. \tag{1.60}$$

We usually write $(\mathcal{L}_{\boldsymbol{X}}\boldsymbol{Y})^\mu$ the components of the new tangent vector as $\mathcal{L}_XY^\mu$. Using the Jacobi identity, one can show that

$$\mathcal{L}_{\boldsymbol{X}}\mathcal{L}_{\boldsymbol{Y}}\boldsymbol{Z}-\mathcal{L}_{\boldsymbol{Y}}\mathcal{L}_{\boldsymbol{X}}\boldsymbol{Z}=\mathcal{L}_{[\boldsymbol{X},\boldsymbol{Y}]}\boldsymbol{Z}. \tag{1.61}$$

The definition of the Lie derivative is extended to one-forms and tensors. One just has to take into account all possible configurations to permute a certain number of objects and remember that a normal sum "lower index with an upper index" creates a plus sign and "an upper index with a lower index" creates a minus sign. For a $(1,1)$ tensor $\boldsymbol{T}$, the Lie derivative along $\boldsymbol{X}$ reads

$$\mathcal{L}_X T^\mu{}_\nu = X^\tau \partial_\tau T^\mu{}_\nu + T^\mu{}_\tau \partial_\nu X^\tau - T^\tau{}_\nu \partial_\tau X^\mu. \tag{1.62}$$

Example. • In the two dimensional plane in Cartesian coordinates, let us consider $f(x,y) = x^2 - \sin y$ and $\boldsymbol{X} = \sin x \partial_y - y^2 \partial_x$. Then, the Lie derivative of f along $\boldsymbol{X}$ is

$$\mathcal{L}_X f = (\sin x \partial_y - y^2 \partial_x)(x^2 - \sin y) = -\sin x \cos y - 2xy^2. \tag{1.63}$$

• In the two dimensional plane in Cartesian coordinates, let us consider $\boldsymbol{\omega} = y^2 \mathrm{d}x + x^2 \mathrm{d}y$ and $\boldsymbol{X} = \partial_x + xy\partial_y$. Then, the Lie derivative components of $\boldsymbol{\omega}$ along $\boldsymbol{X}$ is

$$\begin{aligned} \mathcal{L}_X \omega_0 &= X^\tau \partial_\tau \omega_0 + \omega_\tau \partial_0 X^\tau = (\partial_x + xy\partial_y) y^2 + y^2 \partial_x 1 + x^2 \partial_x xy = 2xy^2 + x^2 y \\ \mathcal{L}_X \omega_1 &= X^\tau \partial_\tau \omega_1 + \omega_\tau \partial_1 X^\tau = 2x + x^3. \end{aligned} \tag{1.64}$$

Combining the two results, we can write

$$\mathcal{L}_{\boldsymbol{X}} \boldsymbol{\omega} = (2xy^2 + x^2 y)\mathrm{d}x + (2x + x^3)\mathrm{d}y. \tag{1.65}$$

Given two tangent vector fields $\boldsymbol{X}$ and $\boldsymbol{Y}$, one can define the commutator $[\boldsymbol{X}, \boldsymbol{Y}] = \boldsymbol{X}\boldsymbol{Y} - \boldsymbol{Y}\boldsymbol{X}$ that can be written in terms of components in the following way

$$[\boldsymbol{X}, \boldsymbol{Y}]^\nu = \left(X^\mu \frac{\partial Y^\nu}{\partial x^\mu} - Y^\mu \frac{\partial X^\nu}{\partial x^\mu} \right). \tag{1.66}$$

The commutator satisfies the Jacobi identity

$$[\boldsymbol{X}, [\boldsymbol{Y}, \boldsymbol{Z}]] + [\boldsymbol{Y}, [\boldsymbol{Z}, \boldsymbol{X}]] + [\boldsymbol{Z}, [\boldsymbol{X}, \boldsymbol{Y}]] = 0, \tag{1.67}$$

The commutator is also a tangent vector and so acts on functions to give a number.

The Lie derivative of a function f along $\boldsymbol{X}$ is

$$\mathcal{L}_{\boldsymbol{X}} f = \boldsymbol{X}(f) = X^\mu \frac{\partial f}{\partial x^\mu}. \tag{1.68}$$

The Lie derivative of a tangent vector field $\boldsymbol{Y}$ along $\boldsymbol{X}$ is the commutator

$$\mathcal{L}_{\boldsymbol{X}} \boldsymbol{Y} = [\boldsymbol{X}, \boldsymbol{Y}]. \tag{1.69}$$

The Lie derivative satisfies

$$\mathcal{L}_{\boldsymbol{X}} \mathcal{L}_{\boldsymbol{Y}} \boldsymbol{Z} - \mathcal{L}_{\boldsymbol{Y}} \mathcal{L}_{\boldsymbol{X}} \boldsymbol{Z} = \mathcal{L}_{[\boldsymbol{X}, \boldsymbol{Y}]} \boldsymbol{Z}, \tag{1.70}$$

and can be extended to one-forms and tensors. For a $(1,1)$ tensor $\boldsymbol{T}$, the Lie derivative along $\boldsymbol{X}$ reads[a]

$$\mathcal{L}_X T^\mu{}_\nu = X^\tau \partial_\tau T^\mu{}_\nu + T^\mu{}_\tau \partial_\nu X^\tau - T^\tau{}_\nu \partial_\tau X^\mu. \tag{1.71}$$

[a] We should remember that a $()_\mu()^\mu$ gives a "+" and $()^\mu()_\mu$ gives a "-".

Chapter 2

Riemannian Geometry

2.1 Abstract and Component Geometry

At this stage, you might wonder why things are defined in a very abstract way when in practice we only use components to explicitly write expressions. Let us introduce two different aspects of geometry which in fact are useful in different situations.

Abstract differential geometry treats a tangent vector as existing in its own right, without necessity to give its breakdown into components

$$\boldsymbol{X} = X^\mu \partial_\mu = X^0 \partial_0 + X^1 \partial_1 + ..., \tag{2.1}$$

just as one is accustomed nowadays in electromagnetism to treat the electric field $\boldsymbol{E}$, without having to write out its components. The abstract approach is useful when one wants to derive results in a simple way. For example in electromagnetism, considering the gradient operator $\boldsymbol{\nabla}$ instead of its components ∂_x, ∂_y, etc is the quickest and simplest mathematical scheme one knows to derive general results.

Differential geometry as expressed in the language of **components** is convenient or necessary or both when one is dealing even at the level of elementary algebra with the most simple applications of Relativity, from the expression of the Friedmann Universe to the curvature around a static black hole. In practice, one uses differential geometry as expressed in terms of components. But never forget that the mathematics of General Relativity can be seen as abstract objects existing in their own right. The last picture is suitable for generalizing General Relativity for example when working in quantum gravity.

Example. Let us take the example of the Lie derivative of a tangent vector $\boldsymbol{Y}$ along $\boldsymbol{X}$. One can either define it by

$$\mathcal{L}_{\boldsymbol{X}} \boldsymbol{Y} = [\boldsymbol{X}, \boldsymbol{Y}], \tag{2.2}$$

or by

$$\mathcal{L}_X Y^\mu = \left(X^\mu \frac{\partial Y^\nu}{\partial x^\mu} - Y^\mu \frac{\partial X^\nu}{\partial x^\mu} \right). \tag{2.3}$$

However note that the Jacobi identity for the Lie derivative is easier to derive using abstract notation than in terms of components.

2.2 Metric

We have yet to meet the star of the show. There is one object that we can place on a manifold whose importance dwarfs all others, at least when it comes to understanding gravity. This is the metric. The existence of a metric brings a whole host of new concepts to the table which, collectively, are called **Riemannian geometry**.

We all know that the metric is a way to measure distances between points on a manifold. It does, indeed, provide this service but it is not its initial purpose. Instead, the metric is an inner product on each tangent vector space $\mathcal{T}_p(\mathcal{M})$.

Definition. A **metric** $\boldsymbol{g}$ is a $(0,2)$ tensor field that is:

(i) Symmetric $\boldsymbol{g}(\boldsymbol{X},\boldsymbol{Y}) = \boldsymbol{g}(\boldsymbol{Y},\boldsymbol{X})$.

(ii) Non-degenerate *i.e.* if for any $p \in \mathcal{M}$, $\boldsymbol{g}(\boldsymbol{X},\boldsymbol{Y})|_p = 0$ for all $\boldsymbol{Y} \in \mathcal{T}_p(\mathcal{M})$ then $\boldsymbol{X}_p = 0$.

With a choice of coordinate, one can write the metric as

$$\boldsymbol{g} = g_{\mu\nu}\,\mathrm{d}x^\mu \mathrm{d}x^\nu. \tag{2.4}$$

The object $\boldsymbol{g}$ is often written as a line element $\mathrm{d}s^2$ and this expression is abbreviated as

$$\mathrm{d}s^2 = g_{\mu\nu}\,\mathrm{d}x^\mu \mathrm{d}x^\nu. \tag{2.5}$$

Note that the metric can be seen as a matrix $g_{\mu\nu}$. If so, this matrix is symmetric $g_{\mu\nu} = g_{\nu\mu}$ with real coefficients[1]. Using the spectral theorem, we know that the metric can be diagonalized *i.e.* we can always pick a basis $\boldsymbol{e}_\mu$ of $\mathcal{T}_p(\mathcal{M})$ so that the matrix $g_{\mu\nu}$ is diagonal. The non-degeneracy condition ensures that none of these diagonal elements vanish. Some are positive, some are negative. The number of negative entries is called the **signature** of the metric. In classical General Relativity, one always uses $(-,+,+,+,...)$ like the Minkowsky metric. There is some mild logic behind this particular convention. When thinking about geometry, the choice $(-,+,+,+,...)$ is preferable as it ensures that length distances are positive. When thinking about quantum field theory, the choice $(+,-,-,-,...)$ is preferable as it ensures that frequencies and energies are positive.

In particular, the metric is used to define the norm of a tangent vector $\boldsymbol{X}$

$$||\boldsymbol{X}||^2 = X^\mu X_\mu = g_{\mu\nu}X^\mu X^\nu. \tag{2.6}$$

Here, one can import notions from Special Relativity regarding the terminology of the norm of a vector depending on its sign. Indeed for a tangent vector $\boldsymbol{X}$ describing

[1] Real coefficients come from the fact that the metric represents a physical object: the gravitational field.

an object's velocity along a curve $x^\mu(\lambda)$ parameterized by λ, depending on the sign of $X^\mu X_\mu = g_{\mu\nu}X^\mu X^\nu$, the tangent vector can be spacelike[2], null[3] or timelike[4]

$$\begin{aligned} X^\mu X_\mu &> 0 \qquad \text{spacelike} \\ X^\mu X_\mu &= 0 \qquad \text{null} \\ X^\mu X_\mu &< 0 \qquad \text{timelike} \end{aligned} \tag{2.7}$$

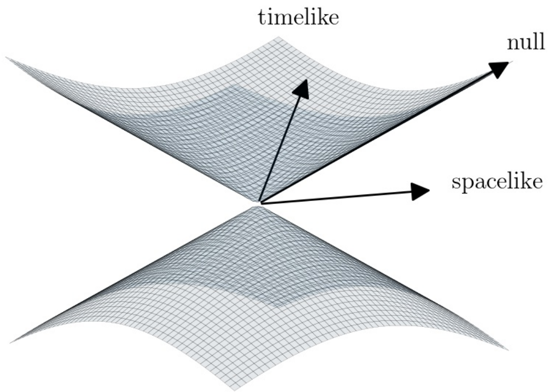


Figure 2.1. Illustration of the light cone and the three types of tangent vectors.

We can use the metric to determine the length of a curve. Given a parametrisation $x^\mu(\lambda)$ of curve with tangent vector $X^\mu = \frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda}$, its length between two points a and b is given by

$$\ell_{a\to b} = \int_a^b d\lambda\sqrt{-g_{\mu\nu}X^\mu X^\nu}. \tag{2.8}$$

Note that for timelike tangent vectors describing most of the physical trajectories, we have $g_{\mu\nu}X^\mu X^\nu < 0$ so the minus sign ensures that we have a positive quantity inside the square root.

Raising and Lowering of Indices. Given a tensor, one can raise and lower indices using the metric. These operations can of course be combined in various ways. For example, given a tangent vector field X^μ, we can associate its dual covector field X_μ

$$X_\mu = g_{\mu\nu}X^\nu, \tag{2.9}$$

and likwise for covectors

$$X^\mu = g^{\mu\nu}X_\nu, \tag{2.10}$$

[2]These vectors describe the velocity of objects travelling faster than the speed of light *i.e.* tachyons.

[3]These vectors describe the velocity of massless objects travelling at the speed of light *i.e.* mainly photons (maybe also gravitons?).

[4]These vectors describe the velocity of massive objects/particles travelling slower than the speed of light.

Note that there are different ways of lowering the indices, and they will in general give rise to different tensors. It is therefore important to keep track of this in the notation. For example $g_{\mu\nu}T^{\nu}{}_{\tau} = T_{\mu\tau}$ and not $T_{\tau\mu}$. This is the reason why the lower indices of a tensor are written more to the right that the upper indices.

Finally note that this notation of raising and lowering indices with the metric is consistent with denoting the inverse metric by raised indices because

$$g^{\mu\nu} = g^{\mu\tau}g^{\nu\rho}g_{\tau\rho}, \tag{2.11}$$

and raising one index of the metric gives the Kronecker tensor,

$$g^{\mu\lambda}g_{\lambda\nu} = g^{\mu}_{\nu} = \delta^{\mu}_{\nu}. \tag{2.12}$$

The metric $\boldsymbol{g}$ is a (0, 2) tensor that enables us to define a line element

$$\mathrm{d}s^2 = \boldsymbol{g} = g_{\mu\nu}\,\mathrm{d}x^{\mu}\mathrm{d}x^{\nu}. \tag{2.13}$$

The metric is symmetric $g_{\mu\nu} = g_{\nu\mu}$ and can always be diagonalized. Its inverse is $g^{\mu\nu}$ so that $g^{\mu\lambda}g_{\lambda\nu} = \delta^{\mu}_{\nu}$. It is also used to define the norm of a tangent vector $\boldsymbol{X}$

$$||\boldsymbol{X}||^2 = X^{\mu}X_{\mu} = g_{\mu\nu}X^{\mu}X^{\nu}. \tag{2.14}$$

When regarding a tangent vector field $\boldsymbol{X}$ as a particle/object's velocity along a trajectory/curve $x^{\mu}(\lambda)$ parametrized with λ, its norm is said to be spacelike, null or timelike depending on its sign. Along such a curve, the metric is used to determine its length between two points a and b

$$\ell_{a\to b} = \int_a^b d\lambda\sqrt{-g_{\mu\nu}X^{\mu}X^{\nu}}. \tag{2.15}$$

2.3 Covariant Derivative and Connection

We have already met one version of differentiation. A tangent vector field $\boldsymbol{X}$ is, at heart, a differential operator and provides a way to differentiate a function f. We write this simply as $\boldsymbol{X}(f)$. As we saw previously, differentiating higher tensor fields is a little more tricky because it requires us to subtract tensor fields at different points. Yet tensors evaluated at different points live in different tangent vector spaces, and it only makes sense to subtract these objects if we can first find a way to map one tangent vector space into the other. To do this, we used the flow generated by a tangent vector $\boldsymbol{X}$ as a way to perform this mapping, resulting in the idea of the Lie derivative $\mathcal{L}_X$.

There is a different way to take derivatives, one which ultimately will prove more useful. The derivative is again associated to a tangent vector field $\boldsymbol{X}$. However, this time we introduce a different object, known as **connection** to map the tangent vector spaces at one point to another tangent vector space at another. The result is an object, distinct from the Lie derivative, called the **covariant derivative**.

2.3.1 Definition of the Connection

First, we give an abstract definition of the covariant derivative, also called connection. This definition will be useful to understand that there are many such connections and so covariant derivatives. Ultimately, we will chose one connection, the Levi-Civita connection, to perform computations in the physical spacetime.

Definition. A **connection** is a map from one tangent vector space $\mathcal{T}_p(\mathcal{M})$ to another $\mathcal{T}_q(\mathcal{M})$. We usually write this map as $\nabla(\boldsymbol{X}, \boldsymbol{Y}) = \nabla_{\boldsymbol{X}}\boldsymbol{Y}$ and the object $\nabla_{\boldsymbol{X}}$ is called the **covariant derivative**. Note that the connection takes as inputs two tangent vectors. It satisfies the following properties for all tangent vectors fields $\boldsymbol{X}, \boldsymbol{Y}$ and $\boldsymbol{Z}$,

(i) $\nabla_{\boldsymbol{X}}(\boldsymbol{Y} + \boldsymbol{Z}) = \nabla_{\boldsymbol{X}}\boldsymbol{Y} + \nabla_{\boldsymbol{X}}\boldsymbol{Z}$.

(ii) $\nabla_{f\boldsymbol{X}+g\boldsymbol{Y}}\boldsymbol{Z} = f\nabla_{\boldsymbol{X}}\boldsymbol{Z} + g\nabla_{\boldsymbol{Y}}\boldsymbol{Z}$ for all functions f and g.

(iii) $\nabla_{\boldsymbol{X}}(f\boldsymbol{Y}) = f\nabla_{\boldsymbol{X}}\boldsymbol{Y} + (\nabla_{\boldsymbol{X}}f)\boldsymbol{Y}$ where we define $\nabla_{\boldsymbol{X}}f = \boldsymbol{X}(f)$.

The covariant derivative endows the manifold $\mathcal{M}$ with more structure. To elucidate this, we can evaluate the connection in a basis ∂_μ of the tangent vector fields. We can always express this as

$$\nabla_{\partial_\rho}\partial_\nu = \Gamma^\mu_{\rho\nu}\partial_\mu, \tag{2.16}$$

with $\Gamma^\mu_{\rho\nu}$ the components of the connection. It is no coincidence that these are denoted by the same greek letter that we use for the Christoffel symbols. However, for now, you should not conflate the two; we will see the relation between them when we introduce the Levi-Civita connection.

Mapping two different tangent vector spaces is what allows the connection to act as a derivative. In what follows, we will use the notation

$$\nabla_\mu = \nabla_{\partial_\mu}. \tag{2.17}$$

This makes the covariant derivative ∇_μ look similar to a partial derivative. Using the properties of the connection, we can write a general derivative of a tangent vector field as

$$\begin{aligned}\nabla_{\boldsymbol{X}}\boldsymbol{Y} &= \nabla_{\boldsymbol{X}}(Y^\mu\partial_\mu) = \boldsymbol{X}(Y^\mu)\partial_\mu + Y^\mu\nabla_{\boldsymbol{X}}\partial_\mu \\ &= X^\nu\partial_\mu Y^\mu\partial_\mu + X^\nu Y^\mu\nabla_\nu\partial_\mu = X^\nu\left(\partial_\nu Y^\mu + \Gamma^\mu_{\nu\rho}Y^\rho\right)\partial_\mu.\end{aligned} \tag{2.18}$$

The fact that we can strip of the overall factor of X^ν means that it makes sense to write the components of the covariant derivative as

$$(\nabla_\nu\boldsymbol{X})^\mu = \nabla_\nu X^\mu = \partial_\nu Y^\mu + \Gamma^\mu_{\nu\rho}Y^\rho, \tag{2.19}$$

where we use the sloppy and sometimes confusing notation $(\nabla_\nu\boldsymbol{X})^\mu = \nabla_\nu X^\mu$. Note that the covariant derivatives coincides with the Lie derivative on functions $\mathcal{L}_{\boldsymbol{X}}f = \nabla_{\boldsymbol{X}}f = \boldsymbol{X}(f)$. It also coincides with the old-fashioned partial derivative $\nabla_\mu f = \partial_\mu f$. However, its action on tangent vector fields differs. In particular, the Lie derivative $\mathcal{L}_{\boldsymbol{X}}\boldsymbol{Y} = [\boldsymbol{X}, \boldsymbol{Y}]$ depends on both $\boldsymbol{X}$ and the first derivative of $\boldsymbol{X}$ while, as we have seen above, the covariant derivative depends only on $\boldsymbol{X}$. This is the property that allows us to write $\nabla_{\boldsymbol{X}} = X^\nu \nabla_\nu$ and think ∇_μ as an operator in its own right. In contrast, there is no way to write "$\mathcal{L}_{\boldsymbol{X}} = X^\mu \mathcal{L}_\mu$".

The Connection is Not a Tensor. We can see this immediately[5] from the definition $\nabla(\boldsymbol{X}, f\boldsymbol{Y}) = \nabla_{\boldsymbol{X}}(f\boldsymbol{Y}) = f\nabla_{\boldsymbol{X}}\boldsymbol{Y} + \boldsymbol{Y}\boldsymbol{X}(f)$. This is not linear in the second argument, which is one of the requirement of a tensor.

However, let us illustrate this using components. We ask what the connection looks like in a different basis

$$\partial'_\mu = \frac{\partial}{\partial x'^\mu} = \frac{\partial x^\nu}{\partial x'^\mu}\frac{\partial}{\partial x^\nu} = J^\nu_\mu \partial_\mu. \tag{2.20}$$

In the basis $\{\partial'_\mu\}$, the connection is written $\nabla_{\partial'_\rho}\partial'_\nu = \Gamma'^\mu_{\rho\nu}\partial'_\mu$. Substituting in the transformation (2.20), we have

$$\Gamma'^\mu_{\rho\nu}\partial'_\mu = \nabla_{J^\sigma_\rho\partial_\sigma}(J^\lambda_\nu\partial_\lambda) = J^\sigma_\rho\nabla_{\partial_\sigma}(J^\lambda_\nu\partial_\lambda) = J^\sigma_\rho J^\lambda_\nu \Gamma^\tau_{\sigma\lambda}\partial_\tau + J^\sigma_{\sigma\rho}\partial_\lambda\partial_\sigma J^\lambda_\nu. \tag{2.21}$$

We can write this as

$$\Gamma'^\mu_{\rho\nu}\partial'_\mu = \left(J^\sigma_\rho J^\lambda_\nu\Gamma^\tau_{\sigma\lambda} + J^\sigma_\rho\partial_\sigma J^\tau_\nu\right)\partial_\tau = \left(J^\sigma_\rho J^\lambda_\nu\Gamma^\tau_{\sigma\lambda} + J^\sigma_\rho\partial_\sigma J^\tau_\nu\right)(J^{-1})^\mu_\tau\partial'_\mu. \tag{2.22}$$

Stripping off the basis tangent vectors ∂'_μ, we see that the components of the connection transform as

$$\Gamma'^\mu_{\rho\nu} = (J^{-1})^\mu_\tau J^\sigma_\rho J^\lambda_\nu\Gamma^\tau_{\sigma\lambda} + (J^{-1})^\mu_\tau J^\sigma_\rho\partial_\sigma J^\tau_\nu. \tag{2.23}$$

The first term coincides with the transformation of a tensor. But the second term, which is independent of Γ, instead depends on ∂J, is novel. This is the characteristic transformation property of a connection.

Differentiating other Tensors. One can use the Leibnizarity of the covariant derivative to extend its action to any tensor field. Here, we give some examples. For a one-form $\boldsymbol{\omega}$, we obtain

$$\nabla_\mu\omega_\nu = \partial_\mu\omega_\nu - \Gamma^\rho_{\mu\nu}\omega_\rho. \tag{2.24}$$

For a $(1,1)$ tensor $\boldsymbol{T}$, one obtains

$$\nabla_\mu T^\nu_\rho = \partial_\mu T^\nu_\rho + \Gamma^\nu_{\mu\tau}T^\tau_\rho - \Gamma^\tau_{\mu\rho}T^\nu_\tau. \tag{2.25}$$

The pattern is clear; for every upper index we get a $+\Gamma T$ term while for every lower index we get a $-\Gamma T$ term.

[5]We see here in practice the power of using abstract differential geometry rather than in terms of components.

The connection $\nabla(\boldsymbol{X},\boldsymbol{Y}) = \nabla_{\boldsymbol{X}}\boldsymbol{Y}$ is a map from one tangent vector space $\boldsymbol{X} \in \mathcal{T}_p(\mathcal{M})$ to another $\boldsymbol{Y} \in \mathcal{T}_q(\mathcal{M})$. The connection is not a tensor. Evaluated in a basis $\{\partial_\mu\}$, it reads

$$\nabla_{\partial_\mu}\partial_\nu = \nabla_\mu\partial_\nu = \Gamma^\sigma_{\mu\nu}\partial_\sigma, \tag{2.26}$$

where $\Gamma^\mu_{\rho\nu}$ are the components of the connection[a]. In terms of components, the connection (or the covariant derivative) reads[b]

$$\nabla_\mu X^\nu = \partial_\mu X^\nu + \Gamma^\nu_{\mu\sigma}X^\sigma. \tag{2.27}$$

The covariant derivative coincides with the Lie derivative for scalar functions $\nabla_{\boldsymbol{X}} f = X^\mu\nabla_\mu f = \mathcal{L}_X f = \boldsymbol{X}(f) = X^\mu\partial_\mu f$. Using the Leibnitz rule, one can extend the action of the covariant derivative to any tensor fields[c]

$$\nabla_\mu\omega_\nu = \partial_\mu\omega_\nu - \Gamma^\rho_{\mu\nu}\omega_\rho. \tag{2.28}$$

$$\nabla_\mu T^\nu_\rho = \partial_\mu T^\nu_\rho + \Gamma^\nu_{\mu\tau}T^\tau_\rho - \Gamma^\tau_{\mu\rho}T^\nu_\tau. \tag{2.29}$$

[a]At this stage, the Γ's are distinct from the Christoffel symbols and so one can define many different connections.

[b]Note that we always use the sometimes confusing notation $\nabla_\mu X^\nu = (\nabla_\mu \boldsymbol{X})^\nu = (\nabla_{\partial_\mu}\boldsymbol{X})^\nu$.

[c]For every upper index we get a $+\Gamma T$ term while for every lower index we get a $-\Gamma T$ term.

2.3.2 Levi-Civita Connection and Christoffel Symbols

So far, note that the connection ∇ has been defined without the use of the metric and that one can define many different connections, depending on the choice for the Γ's. However, something nice happens if we have both a connection and a metric. This nice property is called the **fundamental theorem of Riemannian geometry**.

Theorem. There exists a unique connection that is compatible with the metric $\boldsymbol{g}$ in the sense that

$$\nabla_{\boldsymbol{X}}\boldsymbol{g} = 0 \quad \text{for all tangent vector fields } \boldsymbol{X}. \tag{2.30}$$

This connection is said to be **metric compatible**. In other words, a connection is metric compatible if the covariant derivative of the metric with respect to this connection is everywhere zero. We can demonstrate both the existence and uniqueness of a metric compatible connection by deriving a manifestly unique expression for the connection coefficients in terms of the metric. To accomplish this, we expand out the equation of the metric compatibility for the three different permutations[6] of the indices

$$\begin{cases} \nabla_\sigma g_{\mu\nu} = \partial_\sigma g_{\mu\nu} - \Gamma^\lambda_{\sigma\mu}g_{\lambda\nu} - \Gamma^\lambda_{\sigma\nu}g_{\mu\lambda} = 0 \\ \nabla_\mu g_{\nu\sigma} = \partial_\mu g_{\nu\sigma} - \Gamma^\lambda_{\mu\nu}g_{\lambda\sigma} - \Gamma^\lambda_{\mu\sigma}g_{\nu\lambda} = 0 \\ \nabla_\nu g_{\sigma\mu} = \partial_\nu g_{\sigma\nu} - \Gamma^\lambda_{\nu\sigma}g_{\lambda\mu} - \Gamma^\lambda_{\nu\mu}g_{\sigma\lambda} = 0. \end{cases} \tag{2.31}$$

[6]The metric is symmetric so we have three permutations instead of six.

We subtract the second and third of these from the first, then multiply by the inverse of the metric to obtain[7]

$$\Gamma^{\sigma}_{\mu\nu} = \frac{1}{2} g^{\sigma\rho} \left(\partial_\mu g_{\nu\rho} + \partial_\nu g_{\mu\rho} - \partial_\rho g_{\mu\nu} \right). \tag{2.32}$$

This is one of the most important formulas in this subject; commit it to memory. This connection we have derived from the metric is the one on which conventional General Relativity is based. It is known as the **Levi-Civita connection**. The associated connection coefficients are called **Christoffel symbols**. We then found a link between two independent objects: the metric and the connection. Both concepts can be related if one assumes a particular connection: the Levi-Civita connection.

The fundamental theorem of Riemannian geometry states the existence of a unique connection that is compatible[a] with the metric in the sense that

$$\nabla_\sigma g_{\mu\nu} = 0. \tag{2.33}$$

This connection is called the Levi-Civita connection and its components are the Christoffel symbols

$$\Gamma^{\sigma}_{\mu\nu} = \frac{1}{2} g^{\sigma\rho} \left(\partial_\mu g_{\nu\rho} + \partial_\nu g_{\mu\rho} - \partial_\rho g_{\mu\nu} \right). \tag{2.34}$$

The metric and the connection, two *a priori* independent objects, can be linked by choosing a specific connection: the Levi-Civita one, on which conventional General Relativity is based.

[a]Note that we need also the torsion-free property to define a unique metric compatible connection.

2.3.3 Useful Properties of the Levi-Civita Connection

After defining the Levi-Civita connection, one can rewrite the definition of the Lie derivative by using the covariant derivative. Indeed, the Lie derivative of $\boldsymbol{Y}$ along $\boldsymbol{X}$ is

$$\begin{aligned} \mathcal{L}_X Y^\mu &= X^\nu \partial_\nu Y^\mu - Y^\nu \partial_\nu X^\mu \\ &= X^\nu \nabla_\nu Y^\mu - Y^\nu \nabla_\nu X^\mu - \left(\Gamma^\mu_{\nu\sigma} X^\nu Y^\sigma - \Gamma^\mu_{\nu\sigma} X^\sigma Y^\nu \right) \\ &= X^\nu \nabla_\nu Y^\mu - Y^\nu \nabla_\nu X^\mu, \end{aligned} \tag{2.35}$$

because using the Levi-Civita connection[8], we have $\Gamma^\sigma_{\mu\nu} = \Gamma^\sigma_{\nu\mu}$. This shows that the Lie derivative $\mathcal{L}_X Y^\mu$ is also a tangent vector field that transforms as a tensor. Note, however, that the Lie derivative, in contrast to the covariant derivative, is defined without reference to any metric.

Using the Levi-Civita connection, it is easy to show that the inverse metric also has zero covariant derivative

[7]Note that we obtain the result if we consider the connection coefficients being symmetric with respect to both lower indices. This is a property of a torsion-free manifold. This notion will be introduce later when we define the torsion tensor.

[8]Meaning we also consider the torsion-free property.

$$\nabla_\sigma g_{\mu\nu} = 0 \quad \text{and} \quad \nabla_\sigma g^{\mu\nu} = 0. \tag{2.36}$$

This nice property allows us to raise and lower indices inside the covariant derivative by passing the metric through the covariant derivative. For example, if X_μ is a obtained by lowering an index of the tangent vector X^μ by $X_\mu = g_{\mu\nu}X^\nu$, then

$$\nabla_\sigma X_\mu = \nabla_\sigma(g_{\mu\nu}X^\nu) = g_{\mu\nu}\nabla_\sigma X^\nu, \tag{2.37}$$

where we used both the Leibnitz rule and the fact that $\nabla_\sigma g_{\mu\nu} = 0$. For a higher order tensor, it reads

$$\nabla_\sigma T^\mu_\nu = g^{\mu\tau}\nabla_\sigma T_{\tau\nu}. \tag{2.38}$$

Covariant derivatives commute on scalars. This is of course a familiar property of the ordinary partial derivative, but it is also true for the second covariant derivatives of a scalar and is a consequence of the symmetry of the Christoffel symbols in the second and third indices $\Gamma^\sigma_{\mu\nu} = \Gamma^\sigma_{\nu\mu}$ and is also known as the no torsion property of the covariant derivative. Namely, we have

$$\begin{aligned}\nabla_\mu\nabla_\nu f - \nabla_\nu\nabla_\mu f &= \nabla_\mu\partial_\nu f - \nabla_\nu\partial_\mu f \\ &= \partial_\mu\partial_\nu f - \Gamma^\lambda_{\mu\nu}\partial_\lambda f - \partial_\nu\partial_\mu f - \Gamma^\lambda_{\nu\mu}\partial_\lambda f = 0,\end{aligned} \tag{2.39}$$

because the usual partial derivatives commute. Note that the second covariant derivatives on higher rank tensors do not commute. We will come back to this in our discussion of the curvature tensor later on.

Another useful property is when the covariant derivative lies in between a defined constant norm. Let us consider a tangent vector $\boldsymbol{X}$ with constant norm[9] $X_\mu X^\mu = \epsilon$. Then,

$$\begin{aligned}X^\mu\nabla_\nu X_\mu &= X^\mu\partial_\nu X_\mu - \Gamma^\sigma_{\nu\mu}X^\mu X_\sigma \\ &= \partial_\nu(X_\mu X^\mu) - X_\mu\partial_\nu X^\mu - \Gamma^\sigma_{\nu\mu}X^\mu X_\sigma \\ &= -X_\mu\partial_\nu X^\mu - X_\sigma(\nabla_\nu X^\sigma - \partial_\nu X^\sigma) \\ &= -X_\mu\nabla_\nu X^\mu.\end{aligned} \tag{2.40}$$

Using the fact that one can raise and lower indices inside the covariant derivative with the metric and using the previous result, one obtains

$$X^\mu\nabla_\nu X_\mu = -g_{\mu\alpha}g^{\mu\beta}X^\alpha\nabla_\nu X_\beta = -\delta^\beta_\alpha X^\alpha\nabla_\nu X_\beta = -X^\beta\nabla_\nu X_\beta, \tag{2.41}$$

which leads to $X^\mu\nabla_\nu X_\mu = 0$.

[9] For example, this is the case for velocity tangent vectors parametrized with the proper time. In this case we have $\epsilon = 0, \pm 1$ depending on whether the object/particle is massive or not.

The Lie derivative can be expressed in term of the covariant derivative[a]

$$\mathcal{L}_X Y^\mu = X^\nu \nabla_\nu Y^\mu - Y^\nu \nabla_\nu X^\mu. \tag{2.42}$$

One can raise and lower indices inside the covariant derivative by passing the metric through the covariant derivative[b]. For example,

$$\begin{aligned} \nabla_\sigma X_\mu &= g_{\mu\nu} \nabla_\sigma X^\nu \\ \nabla_\sigma T^\mu_\nu &= g^{\mu\tau} \nabla_\sigma T_{\tau\nu}. \end{aligned} \tag{2.43}$$

Covariant derivatives commute[c] on scalars f

$$\nabla_\mu \nabla_\nu f - \nabla_\nu \nabla_\mu f = 0, \tag{2.44}$$

but it is not the case for higher rank tensors. For tangent vectors $\boldsymbol{X}$ with constant norm $X_\mu X^\mu = \epsilon$, one has

$$X^\mu \nabla_\nu X_\mu = 0. \tag{2.45}$$

[a]This property comes from $\Gamma^\sigma_{\mu\nu} = \Gamma^\sigma_{\nu\mu}$.
[b]This property comes from $\nabla_\sigma g_{\mu\nu} = 0$ and $\nabla_\sigma g^{\mu\nu} = 0$.
[c]This property comes from the fact that usual partial derivatives commute.

2.4 Torsion and Curvature

We saw previously that one can define a connection. In practice, we use the Levi-Civita connection whose components can be expressed in terms of the metric. However, the defined connection is not a tensor. Fortunately, one can use the connection to construct two tensors.

2.4.1 Abstract Definition and Components

The first tensor that one can construct is the **torsion**. It is a $(1,2)$ tensor $\boldsymbol{T}$ and so takes as inputs a one-form $\boldsymbol{\omega}$ and two tangent vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$. It reads

$$\boldsymbol{T}(\boldsymbol{\omega}; \boldsymbol{X}, \boldsymbol{Y}) = \boldsymbol{\omega}(\nabla_{\boldsymbol{X}} \boldsymbol{Y} - \nabla_{\boldsymbol{Y}} \boldsymbol{X} - [\boldsymbol{X}, \boldsymbol{Y}]). \tag{2.46}$$

The second tensor that one can construct is the **Riemann tensor**, also called **curvature**. It is a $(1,3)$ tensor $\boldsymbol{R}$ and so takes as inputs a one-form $\boldsymbol{\omega}$ and three tangent vectors $\boldsymbol{X}, \boldsymbol{Y}$ and $\boldsymbol{Z}$. It reads

$$\boldsymbol{R}(\boldsymbol{\omega}; \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}) = \boldsymbol{\omega}(\nabla_{\boldsymbol{X}} \nabla_{\boldsymbol{Y}} \boldsymbol{Z} - \nabla_{\boldsymbol{Y}} \nabla_{\boldsymbol{X}} \boldsymbol{Z} - \nabla_{[\boldsymbol{X},\boldsymbol{Y}]} \boldsymbol{Z}). \tag{2.47}$$

Alternatively, one could think of torsion $\boldsymbol{T}$ as a map between two tangent vector spaces

$$\boldsymbol{T}(\boldsymbol{X}, \boldsymbol{Y}) = \nabla_{\boldsymbol{X}} \boldsymbol{Y} - \nabla_{\boldsymbol{Y}} \boldsymbol{X} - [\boldsymbol{X}, \boldsymbol{Y}]. \tag{2.48}$$

Similarly, the Riemann tensor $\boldsymbol{R}$ can be viewed as a map between two tangent vector spaces that gives a differential operator acting on a tangent vector space

$$\boldsymbol{R}(\boldsymbol{X},\boldsymbol{Y}) = \nabla_{\boldsymbol{X}}\nabla_{\boldsymbol{Y}} - \nabla_{\boldsymbol{Y}}\nabla_{\boldsymbol{X}} - \nabla_{[\boldsymbol{X},\boldsymbol{Y}]}. \tag{2.49}$$

It is not obvious from the definition that these objects are tensors. To actually demonstrate that $\boldsymbol{T}$ and $\boldsymbol{R}$ are tensors, we need to show that they are linear in all arguments. This is a good example of using the abstract notation to derive expressions. Linearity in $\boldsymbol{\omega}$ is straightforward. For the others, there are some small calculations to do. For example, we must show that $\boldsymbol{T}(\boldsymbol{\omega}; f\boldsymbol{X},\boldsymbol{Y}) = f\boldsymbol{T}(\boldsymbol{\omega};\boldsymbol{X},\boldsymbol{Y})$ where f is a scalar function. To see this, we just run through the definition of the various objects

$$\boldsymbol{T}(\boldsymbol{\omega}; f\boldsymbol{X},\boldsymbol{Y}) = \boldsymbol{\omega}(\nabla_{f\boldsymbol{X}}\boldsymbol{Y} - \nabla_{\boldsymbol{Y}}(f\boldsymbol{X}) - [f\boldsymbol{X},\boldsymbol{Y}]). \tag{2.50}$$

We then use $\nabla_{f\boldsymbol{X}}\boldsymbol{Y} = f\nabla_{\boldsymbol{X}}\boldsymbol{Y}$, $\nabla_{\boldsymbol{Y}}(f\boldsymbol{X}) = f\nabla_{\boldsymbol{Y}}\boldsymbol{X} + \boldsymbol{Y}(f)\boldsymbol{X}$ and $[f\boldsymbol{X},\boldsymbol{Y}] = f[\boldsymbol{X},\boldsymbol{Y}] - \boldsymbol{Y}(f)\boldsymbol{X}$. The two $\boldsymbol{Y}(f)\boldsymbol{X}$ terms cancel, leaving us with

$$\boldsymbol{T}(\boldsymbol{\omega}; f\boldsymbol{X},\boldsymbol{Y}) = f\boldsymbol{\omega}(\nabla_{\boldsymbol{X}}\boldsymbol{Y} - \nabla_{\boldsymbol{Y}}\boldsymbol{X} - [\boldsymbol{X},\boldsymbol{Y}]) = f\boldsymbol{T}(\boldsymbol{\omega};\boldsymbol{X},\boldsymbol{Y}). \tag{2.51}$$

Similarly, for the Riemann tensor we have

$$\begin{aligned}
\boldsymbol{R}(\boldsymbol{\omega}; f\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Z}) &= \boldsymbol{\omega}(\nabla_{f\boldsymbol{X}}\nabla_{\boldsymbol{Y}}\boldsymbol{Z} - \nabla_{\boldsymbol{Y}}\nabla_{f\boldsymbol{X}}\boldsymbol{Z} - \nabla_{[f\boldsymbol{X},\boldsymbol{Y}]}\boldsymbol{Z}) \\
&= \boldsymbol{\omega}(f\nabla_{\boldsymbol{X}}\nabla_{\boldsymbol{Y}}\boldsymbol{Z} - \nabla_{\boldsymbol{Y}}(f\nabla_{\boldsymbol{X}}\boldsymbol{Z}) - \nabla_{f[\boldsymbol{X},\boldsymbol{Y}]}\boldsymbol{Z} + \nabla_{\boldsymbol{Y}(f)\boldsymbol{X}}\boldsymbol{Z}) \\
&= \boldsymbol{\omega}(f\nabla_{\boldsymbol{X}}\nabla_{\boldsymbol{Y}}\boldsymbol{Z} - f\nabla_{\boldsymbol{Y}}\nabla_{\boldsymbol{X}}\boldsymbol{Z} - \boldsymbol{Y}(f)\nabla_{\boldsymbol{X}}\boldsymbol{Z} - f\nabla_{[\boldsymbol{X},\boldsymbol{Y}]}\boldsymbol{Z} + \nabla_{\boldsymbol{Y}(f)\boldsymbol{X}}\boldsymbol{Z}) \\
&= f\boldsymbol{R}(\boldsymbol{\omega};\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Z}).
\end{aligned} \tag{2.52}$$

Linearity in $\boldsymbol{Y}$ follows from the linearity in $\boldsymbol{X}$ because both tangent vectors play a similar role. But we still need to check linearity in $\boldsymbol{Z}$,

$$\begin{aligned}
\boldsymbol{R}(\boldsymbol{\omega};\boldsymbol{X},\boldsymbol{Y}, f\boldsymbol{Z}) &= \boldsymbol{\omega}(\nabla_{\boldsymbol{X}}\nabla_{\boldsymbol{Y}}(f\boldsymbol{Z}) - \nabla_{\boldsymbol{Y}}\nabla_{\boldsymbol{X}}(f\boldsymbol{Z}) - \nabla_{[\boldsymbol{X},\boldsymbol{Y}]}(f\boldsymbol{Z})) \\
&= \boldsymbol{\omega}(\nabla_{\boldsymbol{X}}(f\nabla_{\boldsymbol{Y}}\boldsymbol{Z} + \boldsymbol{Y}(f)\boldsymbol{Z}) - \nabla_{\boldsymbol{Y}}(f\nabla_{\boldsymbol{X}}\boldsymbol{Z} + \boldsymbol{X}(f)\boldsymbol{Z}) \\
&\quad - f\nabla_{[\boldsymbol{X},\boldsymbol{Y}]}\boldsymbol{Z} - [\boldsymbol{X},\boldsymbol{Y}](f)\boldsymbol{Z}) \\
&= \boldsymbol{\omega}(f\nabla_{\boldsymbol{X}}\nabla_{Y} + \boldsymbol{X}(f)\nabla_{\boldsymbol{Y}}\boldsymbol{Z} + \boldsymbol{Y}(f)\nabla_{\boldsymbol{X}}\boldsymbol{Z} + \boldsymbol{X}(\boldsymbol{Y}(f))\boldsymbol{Z} \\
&\quad - f\nabla_{\boldsymbol{Y}}\nabla_{\boldsymbol{X}}\boldsymbol{Z} - \boldsymbol{Y}(f)\nabla_{\boldsymbol{X}}\boldsymbol{Z} - \boldsymbol{X}(f)\nabla_{\boldsymbol{Y}}\boldsymbol{Z} - \boldsymbol{Y}(\boldsymbol{X}(f))\boldsymbol{Z} \\
&\quad - f\nabla_{[\boldsymbol{X},\boldsymbol{Y}]}\boldsymbol{Z} - [\boldsymbol{X},\boldsymbol{Y}](f)\boldsymbol{Z}) \\
&= f\boldsymbol{R}(\boldsymbol{\omega};\boldsymbol{X},\boldsymbol{Y},\boldsymbol{Z}).
\end{aligned} \tag{2.53}$$

Thus, both torsion and curvature define new tensors on our manifold.

The next step is to evaluate these tensors in a coordinate basis $\{\partial_\mu\}$ with the dual basis $\{\mathrm{d}x^\mu\}$. The components of the torsion are

$$\begin{aligned}
T^\rho_{\mu\nu} &= \boldsymbol{T}(\mathrm{d}x^\rho;\partial_\mu,\partial_\nu) \\
&= \mathrm{d}x^\rho(\nabla_\mu\partial_\nu - \partial_\nu\partial_\mu - [\partial_\mu,\partial_\nu]) \\
&= \mathrm{d}x^\rho(\Gamma^\sigma_{\mu\nu}\partial_\sigma - \Gamma^\sigma_{\nu\mu}\partial_\sigma) \\
&= \Gamma^\rho_{\mu\nu} - \Gamma^\rho_{\nu\mu},
\end{aligned} \tag{2.54}$$

where we used the fact that $\mathrm{d}x^\rho\partial_\sigma = \delta^\rho_\sigma$ and $[\partial_\mu,\partial_\nu] = 0$. We learn that, even though $\Gamma^\rho_{\mu\nu}$ is not a tensor, the anti-symmetric part $\Gamma^\rho_{\mu\nu} - \Gamma^\rho_{\nu\mu}$ does form a tensor. Clearly, the torsion is anti-symmetric in the lower two indices

$$T^{\rho}_{\mu\nu} = T^{\rho}_{\nu\mu}. \tag{2.55}$$

Connections which are symmetric in the lower indices, so $\Gamma^{\rho}_{\mu\nu} = \Gamma^{\rho}_{\nu\mu}$ have zero torsion. Such connections are said to be **torsion-free**. This is the case for the Levi-Civita connection[10].

The components of the Riemann tensor are given by[11]

$$\begin{aligned}
R^{\sigma}_{\rho\mu\nu} &= \mathrm{d}x^{\sigma}(\nabla_{\mu}\nabla_{\nu}\partial_{\rho} - \nabla_{\nu}\nabla_{\mu}\partial_{\rho} - \nabla_{[\partial_{\mu},\partial_{\nu}]}\partial_{\rho}) \\
&= \mathrm{d}x^{\sigma}(\nabla_{\mu}\nabla_{\nu}\partial_{\rho} - \nabla_{\nu}\nabla_{\mu}\partial_{\rho}) \\
&= \mathrm{d}x^{\sigma}(\nabla_{\mu}(\Gamma^{\lambda}_{\nu\rho}\partial_{\lambda}) - \nabla_{\nu}(\Gamma^{\lambda}_{\mu\rho}\partial_{\lambda})) \\
&= \mathrm{d}x^{\sigma}((\partial_{\mu}\Gamma^{\lambda}_{\nu\rho})\partial_{\lambda} + \Gamma^{\lambda}_{\nu\rho}\Gamma^{\tau}_{\mu\lambda}\partial_{\tau} - (\partial_{\nu}\Gamma^{\lambda}_{\mu\rho})\partial_{\lambda} - \Gamma^{\lambda}_{\mu\rho}\Gamma^{\tau}_{\nu\lambda}\partial_{\tau}) \\
&= \partial_{\mu}\Gamma^{\sigma}_{\nu\rho} - \partial_{\nu}\Gamma^{\sigma}_{\mu\rho} + \Gamma^{\lambda}_{\nu\rho}\Gamma^{\sigma}_{\mu\lambda} - \Gamma^{\lambda}_{\mu\rho}\Gamma^{\sigma}_{\nu\lambda}.
\end{aligned} \tag{2.56}$$

Clearly, the Riemann tensor is anti-symmetric in its last two indices

$$R^{\sigma}_{\rho\mu\nu} = -R^{\sigma}_{\rho\nu\mu}. \tag{2.57}$$

Note that it would be quite unpleasant to have to verify the tensorial nature of this expression by explicitly checking its behaviour under coordinate transformations.

Using the connection, one can define the torsion, a $(1,2)$ tensor $\boldsymbol{T}$, by

$$\boldsymbol{T}(\boldsymbol{\omega}; \boldsymbol{X}, \boldsymbol{Y}) = \boldsymbol{\omega}(\nabla_{\boldsymbol{X}}\boldsymbol{Y} - \nabla_{\boldsymbol{Y}}\boldsymbol{X} - [\boldsymbol{X}, \boldsymbol{Y}]), \tag{2.58}$$

and the Riemann tensor, a $(1,3)$ tensor $\boldsymbol{R}$, by

$$\boldsymbol{R}(\boldsymbol{\omega}; \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z}) = \boldsymbol{\omega}(\nabla_{\boldsymbol{X}}\nabla_{\boldsymbol{Y}}\boldsymbol{Z} - \nabla_{\boldsymbol{Y}}\nabla_{\boldsymbol{X}}\boldsymbol{Z} - \nabla_{[\boldsymbol{X},\boldsymbol{Y}]}\boldsymbol{Z}). \tag{2.59}$$

Their components are obtained by evaluating $\boldsymbol{T}$ and $\boldsymbol{R}$ in a coordinate basis $\{\partial_{\mu}\}$ with the dual basis $\{\mathrm{d}x^{\mu}\}$. One obtains

$$\begin{aligned}
T^{\rho}_{\mu\nu} &= \Gamma^{\rho}_{\mu\nu} - \Gamma^{\rho}_{\nu\mu} \\
R^{\sigma}_{\rho\mu\nu} &= \partial_{\mu}\Gamma^{\sigma}_{\nu\rho} - \partial_{\nu}\Gamma^{\sigma}_{\mu\rho} + \Gamma^{\lambda}_{\nu\rho}\Gamma^{\sigma}_{\mu\lambda} - \Gamma^{\lambda}_{\mu\rho}\Gamma^{\sigma}_{\nu\lambda}.
\end{aligned} \tag{2.60}$$

Connections with vanishing torsion are called torsion-free connections. These connections verify

$$\Gamma^{\rho}_{\mu\nu} = \Gamma^{\rho}_{\nu\mu}. \tag{2.61}$$

The Levi-Civita connection is torsion-free.

[10] Actually, the torsion-free property is one requirement to derive the Levi-Civita connection components in terms of the metric. The fundamental theorem of Riemannian geometry states the existence of a unique, torsion-free, connection that is metric compatible.

[11] Note the slightly counter intuitive, but standard ordering of the indices.

2.4.2 Identities and Properties of the Riemann Tensor

There is a closely related calculation in which both the torsion and te Riemann tensors appear. We look at the commutator of covariant derivatives acting on tangent vector fields. One can check that

$$[\nabla_\mu, \nabla_\nu] Z^\sigma = \nabla_\mu \nabla_\nu Z^\sigma - \nabla_\nu \nabla_\mu Z^\sigma = R^\sigma_{\rho\mu\nu} Z^\rho - T^\rho_{\mu\nu} \nabla_\rho Z^\sigma. \tag{2.62}$$

This expression is known as the **Ricci identity**. Note that in practice, we will always work with the Levi-Civita connection so that the torsion is zero. One then obtains

$$[\nabla_\mu, \nabla_\nu] Z^\sigma = \nabla_\mu \nabla_\nu Z^\sigma - \nabla_\nu \nabla_\mu Z^\sigma = R^\sigma_{\rho\mu\nu} Z^\rho. \tag{2.63}$$

It is perhaps surprising that the commutator $[\nabla_\mu, \nabla_\nu]$, which appears to be a differential operator, has an action on vector fields which (in the absence of torsion, at any rate) is a simple multiplicative transformation. The Riemann tensor measures that part of the commutator of covariant derivatives which is proportional to the vector field, while the torsion tensor measures the part which is proportional to the covariant derivative of the vector field; the second derivative does not enter at all. One must remember that the Riemann tensor measures the failure of the covariant derivative to commute.

The extension of the above formula to any higher order tensors follows the usual pattern, with one Riemann tensor contracted with every index. For example,

$$[\nabla_\mu, \nabla_\nu] T^{\alpha\beta} = \nabla_\mu \nabla_\nu T^{\alpha\beta} - \nabla_\nu \nabla_\mu T^{\alpha\beta} = R^\alpha_{\sigma\mu\nu} T^{\sigma\beta} + R^\beta_{\sigma\mu\nu} T^{\alpha\sigma}. \tag{2.64}$$

There are also a number of symmetric properties satisfied by the Riemann tensor when we use the Levi-Civita connection. We will just state but not prove the following results.

If we lower an index on the Riemann tensor and write $R_{\sigma\rho\mu\nu} = g_{\sigma\lambda} R^\lambda_{\rho\mu\nu}$, then the resulting tensor obeys the following symmetry identities

$$\begin{aligned} R_{\sigma\rho\mu\nu} &= -R_{\sigma\rho\nu\mu} \\ R_{\sigma\rho\mu\nu} &= -R_{\rho\sigma\mu\nu} \\ R_{\sigma\rho\mu\nu} &= R_{\mu\nu\sigma\rho}. \end{aligned} \tag{2.65}$$

The Riemann tensor also satisfies the following cyclic permutation relation

$$R_{\sigma\rho\mu\nu} + R_{\sigma\mu\nu\rho} + R_{\sigma\nu\rho\mu} = 0 \tag{2.66}$$

We can now count how many independent components the Riemann tensor really has in dimension 4. The anti-symmetric property of the last two indices implies that the second pair of indices can only take $(3 \times 4)/2 = 6$ independent values. The anti-symmetric property of the first two indices implies the same. The symmetric property implies that the Riemann tensor can be seen as a symmetric 6×6 matrix and thus has $(6 \times 7)/2 = 21$ components. The cyclic property adds an additional constraint so that the Riemann tensor has 20 independent components. Note that in dimension D, the Riemann tensor has $D^2(D^2 - 1)/12$ independent components.

Finally, the Riemann tensor also satisfies the so-called **Bianchi identity**

$$\nabla_\lambda R_{\sigma\rho\mu\nu} + \nabla_\sigma R_{\rho\lambda\mu\nu} + \nabla_\rho R_{\lambda\sigma\mu\nu} = 0. \tag{2.67}$$

Note that for a general connection there would be additional terms involving the torsion tensor. This cyclic identity is closely related to the Jacobi identity for the covariant derivative

$$[[\nabla_\mu, \nabla_\nu], \nabla_\rho] + [[\nabla_\nu, \nabla_\rho], \nabla_\mu] + [[\nabla_\rho, \nabla_\mu], \nabla_\nu] = 0. \tag{2.68}$$

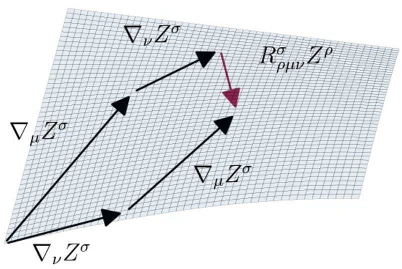


Figure 2.2. Illustration of the Ricci identity with the Levi-Civita connection.

The Riemann tensor measures the failure the failure of the covariant derivative to commute[a]

$$[\nabla_\mu, \nabla_\nu] Z^\sigma = \nabla_\mu \nabla_\nu Z^\sigma - \nabla_\nu \nabla_\mu Z^\sigma = R^\sigma_{\rho\mu\nu} Z^\rho. \tag{2.69}$$

The above formula is extended to any tensor. For example,

$$[\nabla_\mu, \nabla_\nu] T^{\alpha\beta} = \nabla_\mu \nabla_\nu T^{\alpha\beta} - \nabla_\nu \nabla_\mu T^{\alpha\beta} = R^\alpha_{\sigma\mu\nu} T^{\sigma\beta} + R^\beta_{\sigma\mu\nu} T^{\alpha\sigma}. \tag{2.70}$$

The Riemann tensor verifies the following properties

$$\begin{aligned} R_{\sigma\rho\mu\nu} &= -R_{\sigma\rho\nu\mu} \\ R_{\sigma\rho\mu\nu} &= -R_{\rho\sigma\mu\nu} \\ R_{\sigma\rho\mu\nu} &= R_{\mu\nu\sigma\rho} \\ R_{\sigma\rho\mu\nu} &+ R_{\sigma\mu\nu\rho} + R_{\sigma\nu\rho\mu} = 0. \end{aligned} \tag{2.71}$$

and the Bianchi identity

$$\nabla_\lambda R_{\sigma\rho\mu\nu} + \nabla_\sigma R_{\rho\lambda\mu\nu} + \nabla_\rho R_{\lambda\sigma\mu\nu} = 0. \tag{2.72}$$

In dimension D, the Riemann tensor has $D^2(D^2-1)/12$ independent components.

[a] This formula is valid when using a torsion-free connection.

2.5 Ricci and Einstein Tensors

There are a number of further tensors that we can build from the Riemann tensor and that are especially important in General Relativity. First, it is frequently useful to consider contractions of the Riemann tensor. Even without the metric, we can form a contracted Riemann tensor called the **Ricci tensor**

$$R_{\mu\nu} = R^{\sigma}_{\mu\sigma\nu}. \tag{2.73}$$

The Ricci tensor associated with the Levi-Civita connection, which is always used in practice, is symmetric. It inherits its symmetry from the Riemann tensor. We write $R_{\mu\nu} = g^{\sigma\rho} R_{\sigma\mu\rho\nu} = g^{\rho\sigma} R_{\rho\nu\sigma\mu}$, giving

$$R_{\mu\nu} = R_{\nu\mu}. \tag{2.74}$$

One can go one step further with contractions and create a function R over the manifold. This is the **Ricci scalar**, also called **scalar curvature**,

$$R = R^{\mu}_{\mu} = g^{\mu\nu} R_{\mu\nu}. \tag{2.75}$$

This scalar can be seen as the trace of the Ricci tensor. The Bianchi identity has a nice implication for the Ricci tensor. If we write the Bianchi identity

$$\nabla_\lambda R_{\sigma\rho\mu\nu} + \nabla_\sigma R_{\rho\lambda\mu\nu} + \nabla_\rho R_{\lambda\sigma\mu\nu} = 0, \tag{2.76}$$

and multiply by $g^{\mu\lambda} g^{\rho\nu}$, one obtains

$$\nabla^\mu R_{\mu\sigma} - \nabla_\sigma R + \nabla^\nu R_{\nu\sigma} = 0, \tag{2.77}$$

which means that $\nabla^\mu R_{\mu\nu} = \frac{1}{2}\nabla_\nu R$. This motivates us to introduce the **Einstein tensor**

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu}, \tag{2.78}$$

which has the property that it is covariantly constant, or divergent less, meaning

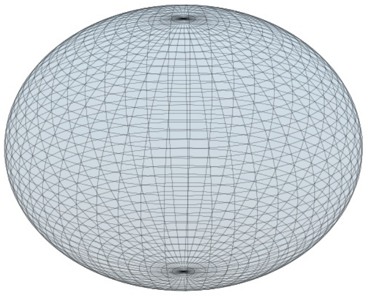
$$\nabla^\mu G_{\mu\nu} = 0. \tag{2.79}$$

The Riemann tensor contains all the information about the curvature of the manifold. However, note that vanishing Christoffel symbols does not mean the manifold is flat. This is because the connection is not a tensor. But the Riemann tensor is a tensor and so if it vanishes in one coordinate system, it must vanish in all of them. Given some horrible coordinate system, with non-vanishing Christoffel symbols, we can always compute the corresponding Riemann tensor and thus the scalar curvature R to see if the manifold is actually flat.

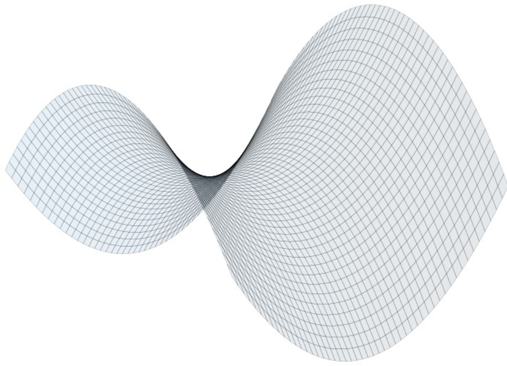
The elementary and universal applicable method for computing the components of the Riemann tensor and thus the Ricci tensor and finally the curvature starts from the metric components in a coordinate basis, and proceeds by the following scheme:

$$g_{\mu\nu} \xrightarrow{\Gamma\sim\partial g} \Gamma^{\sigma}_{\mu\nu} \xrightarrow{R\sim\partial\Gamma+\Gamma\Gamma} R^{\sigma}_{\rho\mu\nu} \xrightarrow{R^{\sigma}_{\mu\sigma\nu}} R_{\mu\nu} \xrightarrow{R^{\mu}_{\mu}} R. \tag{2.80}$$

We see here that if one knows the metric, one knows everything about the manifold. Note that the Christoffel symbols can be obtained in a straightforward by varying the



(a) Positive curvature $R > 0$.



(b) Negative curvature $R < 0$.

Figure 2.3. Surfaces with different curvatures.

action. We will see this in another section.

One obtains the Ricci tensor by contracting the Riemann tensor

$$R_{\mu\nu} = R^{\sigma}_{\mu\sigma\nu}. \tag{2.81}$$

The Ricci scalar, also called scalar curvature, is defined as the trace of the Ricci tensor

$$R = R^{\mu}_{\mu} = g^{\mu\nu} R_{\mu\nu}. \tag{2.82}$$

One can then define the Einstein tensor

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu}, \tag{2.83}$$

which is covariantly constant, or divergent less, meaning

$$\nabla^{\mu} G_{\mu\nu} = 0. \tag{2.84}$$

Starting from the metric, one can have a complete description of the manifold by following this scheme

$$g_{\mu\nu} \xrightarrow{\Gamma\sim\partial g} \Gamma^{\sigma}_{\mu\nu} \xrightarrow{R\sim\partial\Gamma+\Gamma\Gamma} R^{\sigma}_{\rho\mu\nu} \xrightarrow{R^{\sigma}_{\mu\sigma\nu}} R_{\mu\nu} \xrightarrow{R^{\mu}_{\mu}} R. \tag{2.85}$$

2.6 Parallel Transport

Although we have now met a number of properties of the connection and especially the Levi-Civita connection, we have not yet fully explained its name. How does it connect different tangent vector spaces?

The answer is that the connection connects tangent vector spaces at two different points of the the manifold by mean of a map called **parallel transport**. As we stressed earlier, such a map is necessary to define differentiation. Note that it does not make any sense to ask if two tangent vectors are parallel in a curved space. However, given a metric and a curve connecting these two points, one can compare the two by dragging one along the curve to the other using the covariant derivative.

Take a tangent vector field $\boldsymbol{X}$ and consider some associated integral curves with coordinates $x^\mu(\lambda)$, such that

$$X^\mu = \frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda}. \tag{2.86}$$

We say that a tensor field $\boldsymbol{T}$ is **parallely transported** along the defined curve if

$$\nabla_{\boldsymbol{X}} \boldsymbol{T} = 0. \tag{2.87}$$

To illustrate this, consider the parallel transport of a second tangent vector field $\boldsymbol{Y}$. In terms of the components, the last condition reads

$$X^\mu \nabla_\mu Y^\nu = X^\mu (\partial_\mu Y^\nu + \Gamma^\nu_{\mu\rho} Y^\rho). \tag{2.88}$$

If we now evaluate this on the curve thinking of $X^\mu \partial_\mu Y^\nu = \frac{\mathrm{d}Y^\nu}{\mathrm{d}\lambda}$, one obtains

$$\frac{\mathrm{d}Y^\nu}{\mathrm{d}\lambda} + \Gamma^\nu_{\mu\rho} X^\mu Y^\rho = 0. \tag{2.89}$$

These are a set of coupled, ordinary differential equations. Given an initial condition $x^\mu(\lambda_0)$ corresponding to a point $p \in \mathcal{M}$, these equations can be solved to find a unique tangent vector at each point along the curve.

Parallel transport is path dependent. It depends on both the connection and the underlying path wich, in this case, is characterised by the tangent vector field $\boldsymbol{X}$. To parallely transport an object between two points, one must first define the parametrized curve.

Example. On the two dimensional sphere in spherical coordinates (θ, φ), let us parallely transport the vector $\boldsymbol{Y} = \partial_\varphi = (0, 1)$ between the point $p = (\theta = \alpha, \varphi = \varphi_0)$ and $q = (\theta = \alpha, \varphi = \varphi_0 + \delta)$ that is along a curve parallel to the equator. First we need to define a parametrized curve $\mathcal{C}$. Here we take $\mathcal{C} = \{\theta = \alpha, \varphi = \varphi_0 + \delta\, \lambda\}$ where λ is the parameter. We then define the tangent vector corresponding to this integral curve

$$\boldsymbol{X} = X^\mu \partial_\mu = \frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda} \partial_\mu = 0 \times \partial_\theta + \delta \partial_\varphi = (0, \delta), \tag{2.90}$$

so that

$$X^\mu \nabla_\mu Y^\nu = 0 \quad \text{leads to} \quad \nabla_\varphi Y^\mu = 0. \tag{2.91}$$

Using the definition of the covariant derivative in terms of the components, the last equation is a set of coupled first order differential equations

$$\begin{cases} \partial_\varphi Y^\theta + \Gamma^\theta_{\varphi\varphi} Y^\varphi = 0 \\ \partial_\varphi Y^\varphi + \Gamma^\varphi_{\varphi\theta} Y^\theta = 0. \end{cases} \tag{2.92}$$

In practice, these equations are very hard to solve. Here, we know the Christoffel symbols and the fact that θ is constant. Solving these equations (with constant $\theta = \alpha$) where the constants of integration are found with the initial tangent vector leads to

$$\begin{cases} Y^\theta = \sin\alpha \, \sin(\cos\alpha \, \delta) \\ Y^\varphi = \cos(\cos\alpha \, \delta). \end{cases} \tag{2.93}$$

Note that after a round trip ($\delta = 2\pi$), we do not recover the same tangent vector except along the equator $\theta = \alpha = \pi/2$.

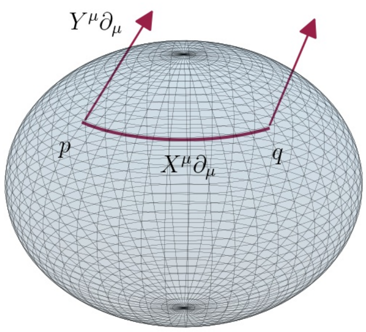


Figure 2.4. Illustration of the above example of parallel transport on a sphere.

The connection maps to tangent vector spaces at two different points of the manifold by the parallel transport. A tensor $\boldsymbol{T}$ is parallely transported along an integral curve $x^\mu(\lambda)$ defined by the tangent vector field $\boldsymbol{X}$ which satisfies $X^\mu = \frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda}$ if

$$\nabla_{\boldsymbol{X}} \boldsymbol{T} = 0. \tag{2.94}$$

For a tangent vector field $\boldsymbol{Y}$, it reads in terms of components

$$\frac{\mathrm{d}Y^\nu}{\mathrm{d}\lambda} + \Gamma^\nu_{\mu\rho} X^\mu Y^\rho = 0. \tag{2.95}$$

These are a set of coupled, ordinary differential equations that can be solved exactly only in a very few cases.

2.7 Geodesics and auto-parallels

A **geodesic** is a curve tangent to a tangent vector field $\boldsymbol{X}$ that obeys

$$\nabla_{\boldsymbol{X}} \boldsymbol{X} = 0. \tag{2.96}$$

Along a curve $x^\mu(\lambda)$ parametrized by λ, we can write the above equation in terms of components

$$\frac{d^2x^\mu}{d\lambda^2} + \Gamma^\mu_{\nu\rho}\frac{dx^\nu}{d\lambda}\frac{dx^\rho}{d\lambda} = 0. \tag{2.97}$$

This is precisely the geodesic equation. We can characterise geodesics by the property that their tangent vectors are parallely transported (do not change) along the curve. For this reason geodesics are also known as **auto-parallels**.

Note that for the Levi-Civita connection, we have $\nabla_{\boldsymbol{X}}\boldsymbol{g} = 0$. This ensures that for any tangent vector field $\boldsymbol{Y}$ parallely transported along a geodesic $\boldsymbol{X}$, we have

$$\frac{d}{d\lambda}\boldsymbol{g}(\boldsymbol{X}, \boldsymbol{Y}) = 0, \tag{2.98}$$

which tells us that $\boldsymbol{Y}$ makes the same angle with the tangent vector $\boldsymbol{X}$ along each point of the geodesic.

A geodesic, also called auto-parallel curve, is a curve that is always tangent to its tangent vector $\boldsymbol{X}$, namely

$$\nabla_{\boldsymbol{X}}\boldsymbol{X} = 0, \tag{2.99}$$

which in terms of components reads

$$\frac{d^2x^\mu}{d\lambda^2} + \Gamma^\mu_{\nu\rho}\frac{dx^\nu}{d\lambda}\frac{dx^\rho}{d\lambda} = 0. \tag{2.100}$$

These are coupled differential equations that can be solved to find the curve parametrized by $x^\mu(\lambda)$.

2.7.1 Affine and Non-affine Parametrisations

To understand the significance of how one parametrises geodesics, observe that the geodesic equation (2.97) is not reparametrisation invariant. Indeed, if one change of parametrisation $\lambda \to \tilde{\lambda}$, then

$$\frac{dx^\mu}{d\lambda} = \frac{d\tilde{\lambda}}{d\lambda}\frac{dx^\mu}{d\tilde{\lambda}}, \tag{2.101}$$

and therefore the geodesic equation can be written

$$\frac{d^2x^\mu}{d\tilde{\lambda}^2} + \Gamma^\mu_{\nu\rho}\frac{dx^\nu}{d\tilde{\lambda}}\frac{dx^\rho}{d\tilde{\lambda}} = -\left(\frac{d\lambda}{d\tilde{\lambda}}\right)^2\frac{d^2\tilde{\lambda}}{d\lambda^2}\frac{dx^\mu}{d\tilde{\lambda}}. \tag{2.102}$$

Thus the geodesic equation retains its form only under affine changes $\tilde{\lambda} = a\lambda + b$ so that $\frac{d^2\tilde{\lambda}}{d\lambda^2} = 0$. Parameters that make the right-hand side of (2.102) vanish are called **affine parameters** and are related to each other by affine transformations. The geodesic equation (2.97) written in this form is said to be affinely parameterised.

Conversely, if we find a curve $x^\mu(\tilde{\lambda})$ that satisfies

$$\frac{\mathrm{d}^2 x^\mu}{\mathrm{d}\tilde{\lambda}^2} + \Gamma^\mu_{\nu\rho}\frac{\mathrm{d}x^\nu}{\mathrm{d}\tilde{\lambda}}\frac{\mathrm{d}x^\rho}{\mathrm{d}\tilde{\lambda}} = C(\tilde{\lambda})\frac{\mathrm{d}x^\mu}{\mathrm{d}\tilde{\lambda}}, \tag{2.103}$$

for some function $C(\tilde{\lambda})$, we can deduce that this curve is the trajectory of a geodesic, but that it is simply not parametrised by an affine parameter. Denoting $f(\tilde{\lambda}) = \lambda$, an affine parameter λ is determined by

$$C(\tilde{\lambda}) = -\frac{\ddot{f}}{\dot{f}^2} \quad i.e. \quad \frac{\mathrm{d}\lambda}{\mathrm{d}\tilde{\lambda}} = \exp\left(\int \mathrm{d}s\, C(s)\right), \tag{2.104}$$

where $\dot{f} = \frac{\mathrm{d}\lambda}{\mathrm{d}\tilde{\lambda}}$. So for $C(\tilde{\lambda}) = 0$, we have an affine transformation between λ and $\tilde{\lambda}$. Note that the proper time τ is an affine parameter for massive particle trajectories.

In general, an auto-parallel curve in terms of components is such that

$$\frac{\mathrm{d}^2 x^\mu}{\mathrm{d}\lambda^2} + \Gamma^\mu_{\nu\rho}\frac{\mathrm{d}x^\nu}{\mathrm{d}\lambda}\frac{\mathrm{d}x^\rho}{\mathrm{d}\lambda} = C(\lambda)\frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda}, \tag{2.105}$$

where $C(\lambda)$ is a function of the parameter λ. However, one can always chose a parameter λ for which $C(\lambda) = 0$. Such parameters are called affine parameters. Affine parameters are related to each other by an affine transformation $\lambda \to \tilde{\lambda} = a\lambda + b$. For example, proper time[a] τ for massive particles is an affine parameter.

[a]Defined by $\mathrm{d}s^2 = -c^2\mathrm{d}^2\tau$.

2.7.2 Geodesics as Extremal Line Elements

If we assume the differential manifold $\mathcal{M}$ to be the four dimensional spacetime, all objects follow geodesics. In other words, objects follow paths that extremizes their line element[12]. Let us take the example of a timelike[13] particle trajectory parametrized by $X^\mu = \frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda}$. Here, X^μ is associated with the particle's velocity. This is why we usually denote this vector $\boldsymbol{V}$ or $\boldsymbol{U}$. The elementary trajectory length is

$$\mathrm{d}s = \sqrt{-g_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu}. \tag{2.106}$$

We then introduce the following action

$$S = \int \mathrm{d}\lambda\, \mathcal{L} = \int \mathrm{d}\lambda \sqrt{-g_{\mu\nu}X^\mu X^\nu}. \tag{2.107}$$

We know that extremizing the action leads to the **Euler-Lagrange equations**

$$\frac{\partial \mathcal{L}}{\partial x^\mu} = \frac{\mathrm{d}}{\mathrm{d}\lambda}\left(\frac{\partial \mathcal{L}}{\partial X^\mu}\right) \quad \text{with} \quad X^\mu = \frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda}. \tag{2.108}$$

Let us now apply the Euler-Lagrange to the previous defined Lagrangian $\mathcal{L} = \sqrt{-g_{\mu\nu}\mathrm{d}X^\mu \mathrm{d}X^\nu} = \sqrt{-g_{\mu\nu}\frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda}\frac{\mathrm{d}x^\nu}{\mathrm{d}\lambda}}$. Recalling that the metric $g_{\mu\nu}$ depends on x^μ, one obtains

[12]This is the principle of least action.
[13]Such that $X_\mu X^\mu = g_{\mu\nu}X^\mu X^\nu < 0$.

$$\begin{aligned}\frac{\partial\mathcal{L}}{\partial x^\sigma} &= -\frac{1}{2\mathcal{L}}\frac{\partial g_{\mu\nu}}{\partial x^\sigma}X^\mu X^\nu \\ \frac{\partial\mathcal{L}}{\partial X^\sigma} &= -\frac{1}{\mathcal{L}}g_{\sigma\nu}X^\nu,\end{aligned} \tag{2.109}$$

so that, using the Leibnitz rule and the chain rule $\frac{\mathrm{d}}{\mathrm{d}\lambda} = X^\alpha\partial_\alpha$, one has

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\left(\frac{\partial\mathcal{L}}{\partial X^\sigma}\right) = g_{\sigma\nu}X^\nu\frac{\mathrm{d}}{\mathrm{d}\lambda}\left(\frac{1}{\mathcal{L}}\right) + \frac{1}{\mathcal{L}}X^\alpha\frac{\partial g_{\sigma\nu}}{\partial x^\alpha}X^\nu + \frac{1}{\mathcal{L}}g_{\sigma\nu}\frac{\mathrm{d}X^\nu}{\mathrm{d}\lambda}. \tag{2.110}$$

Writing down the Euler-Lagrange equation with all the terms and rearranging the expressions leads to

$$g_{\sigma\nu}\frac{\mathrm{d}^2x^\nu}{\mathrm{d}\lambda^2} + \left(\partial_\mu g_{\sigma\nu} - \frac{1}{2}\partial_\sigma g_{\mu\nu}\right)\frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda}\frac{\mathrm{d}x^\nu}{\mathrm{d}\lambda} = C(\lambda)g_{\sigma\nu}\frac{\mathrm{d}x^\nu}{\mathrm{d}\lambda}, \tag{2.111}$$

where we have defined $C(\lambda) = \frac{1}{\mathcal{L}}\frac{\mathrm{d}\mathcal{L}}{\mathrm{d}\lambda}$. A standard trick here is to see that the first term inside the parenthesis can be symmetrized

$$\partial_\mu g_{\sigma\nu} = \frac{1}{2}\left(\partial_\mu g_{\sigma\nu} + \partial_\nu g_{\sigma\mu}\right). \tag{2.112}$$

Multiplying the above equation by the inverse metric $g^{\alpha\sigma}$ and using the definition of the Christoffel symbols, one finally obtains

$$\frac{\mathrm{d}^2x^\alpha}{\mathrm{d}\lambda^2} + \Gamma^\alpha_{\mu\nu}\frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda}\frac{\mathrm{d}x^\nu}{\mathrm{d}\lambda} = C(\lambda)\frac{\mathrm{d}x^\alpha}{\mathrm{d}\lambda}. \tag{2.113}$$

We note that this is the geodesic equation using a non-affine parameter. However, if one executes the same steps with the following Lagrangian

$$\mathcal{L} = -g_{\mu\nu}X^\mu X^\nu, \tag{2.114}$$

the resulting Euler-Lagrange equations leads to an affinely parametrized geodesic

$$\frac{\mathrm{d}^2x^\alpha}{\mathrm{d}\lambda^2} + \Gamma^\alpha_{\mu\nu}\frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda}\frac{\mathrm{d}x^\nu}{\mathrm{d}\lambda} = 0. \tag{2.115}$$

The Euler-Lagrange equation associated with a Lagrangian $\mathcal{L}(x^\mu, \frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda})$ is

$$\frac{\partial \mathcal{L}}{\partial x^\mu} = \frac{\mathrm{d}}{\mathrm{d}\lambda}\left(\frac{\partial \mathcal{L}}{\partial X^\mu}\right) \quad \text{with} \quad X^\mu = \frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda} = \dot{x}^\mu. \tag{2.116}$$

Extremizing the element line with the Lagrangian $\mathcal{L} = \sqrt{-g_{\mu\nu}\frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda}\frac{\mathrm{d}x^\nu}{\mathrm{d}\lambda}}$ leads to a non-affinely parametrized geodesic

$$\frac{\mathrm{d}^2 x^\alpha}{\mathrm{d}\lambda^2} + \Gamma^\alpha_{\mu\nu}\frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda}\frac{\mathrm{d}x^\nu}{\mathrm{d}\lambda} = C(\lambda)\frac{\mathrm{d}x^\alpha}{\mathrm{d}\lambda}, \tag{2.117}$$

where $C(\lambda) = \frac{1}{\mathcal{L}}\frac{\mathrm{d}\mathcal{L}}{\mathrm{d}\lambda}$. However, one can simplify the derivation by considering $\mathcal{L} = -g_{\mu\nu}\frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda}\frac{\mathrm{d}x^\nu}{\mathrm{d}\lambda}$. Extremizing the action with this Lagrangian leads to an affinely parametrized geodesic

$$\frac{\mathrm{d}^2 x^\alpha}{\mathrm{d}\lambda^2} + \Gamma^\alpha_{\mu\nu}\frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda}\frac{\mathrm{d}x^\nu}{\mathrm{d}\lambda} = 0. \tag{2.118}$$

One can always choose an affine parameter (for example proper time for massive particles) and apply the Euler-Lagrange equations without the square root.

2.7.3 Computing Christoffel Symbols from the Action

If the answer to a problem or the result of a computation is not simple, then there is no simple way to obtain it. But when a long computation gives a short answer, then one looks for a better method. It is exactly the case when computing Christoffel symbols from the metric. It usually involves much wasted effort. Indeed in most cases, one computes many Γ's that turn out to be zero. Here, we present the "geodesic Lagrangian" method that provides an economical way to tabulate the Γs.

One normally thinks that the connection coefficients $\Gamma^\sigma_{\mu\nu}$ must be known before one can write the geodesic equation. However, when one affinely parametrizes the geodesic equation, it is simple to apply the Euler-Lagrange equation and explicitly find the geodesic equation. Thus, one can use this fact to compute the Christoffel symbols.

To do this, one must explicitly write down the Lagrangian that leads to an affine parametrised geodesic

$$\mathcal{L} = -g_{\mu\nu}\frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda}\frac{\mathrm{d}x^\nu}{\mathrm{d}\lambda} = -g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu, \tag{2.119}$$

and use the Euler-Lagrange equation. One then identifies the resulting equations with the affinely parametrized geodesic equations and reads out the Christoffel symbols.

Example. Let us consider the two dimensional sphere parametrized with (θ, φ). We want to compute the Christoffel symbols without using the formula involving the metric. Instead, we write the affinely parametrized Lagrangian

$$\mathcal{L} = -g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu = -\left(\dot{\theta}^2 + \sin^2\theta\,\dot{\varphi}^2\right). \tag{2.120}$$

The Euler-Lagrange for $x^0 = \theta$ leads to

$$\frac{\partial \mathcal{L}}{\partial \theta} = -2\sin\theta\cos\theta\,\dot{\varphi}^2, \qquad \frac{\partial \mathcal{L}}{\partial \dot{\theta}} = -2\dot{\theta} \quad \text{and} \quad \frac{\mathrm{d}}{\mathrm{d}\lambda}\frac{\partial \mathcal{L}}{\partial \dot{\theta}} = -2\ddot{\theta}, \tag{2.121}$$

so that

$$\ddot{\theta} - \sin\theta\cos\theta\,\dot{\varphi}^2 = 0. \tag{2.122}$$

The Euler-Lagrange for $x^1 = \varphi$ leads to

$$\frac{\partial \mathcal{L}}{\partial \varphi} = 0, \qquad \frac{\partial \mathcal{L}}{\partial \dot{\varphi}} = -2\sin^2\theta\,\dot{\varphi} \quad \text{and} \quad \frac{\mathrm{d}}{\mathrm{d}\lambda}\frac{\partial \mathcal{L}}{\partial \dot{\varphi}} = -4\sin\theta\cos\theta\dot{\theta}\dot{\varphi} - 2\sin^2\theta\,\ddot{\varphi}, \tag{2.123}$$

so that

$$\ddot{\varphi} + 2\frac{\cos\theta}{\sin\theta}\,\dot{\theta}\dot{\varphi} = 0. \tag{2.124}$$

These two equations enable us to read out the Christoffel symbols

$$\Gamma^{\theta}_{\varphi\varphi} = -\sin\theta\cos\theta \quad \text{and} \quad \Gamma^{\varphi}_{\theta\varphi} = \Gamma^{\theta}_{\varphi\theta} = \frac{\cos\theta}{\sin\theta}. \tag{2.125}$$

One can computes with less effort the connection components $\Gamma^{\sigma}_{\mu\nu}$ by writing down the Lagrangian

$$\mathcal{L} = -g_{\mu\nu}\frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda}\frac{\mathrm{d}x^\nu}{\mathrm{d}\lambda} = -g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu, \tag{2.126}$$

and using the Euler-Lagrange equation. One then identifies the resulting equations with the affinely parametrized geodesic equations and reads out the Christoffel symbols.

Chapter 3

Advanced Topics

3.1 Symmetries

We all know that symmetries are very important in physics because in most cases they simplify the problem. We also know that symmetries are very important because a symmetry implies that something is conserved. This is of course the Nœther's theorem. Here, we discuss the symmetries of the spacetime metric. Intuitively, the notion of symmetry is clear. If you hold up a round sphere, it looks the same no matter what way you rotate it. We want a way to state this mathematically.

3.1.1 Killing Vectors

Let us first recall the expression of the Lie derivative along $\boldsymbol{\xi}$ of a $(0,2)$ tensor $\boldsymbol{T}$ written in terms of the covariant derivative

$$\mathcal{L}_\xi T_{\mu\nu} = \xi^\sigma \nabla_\sigma T_{\mu\nu} + T_{\mu\sigma}\nabla_\nu \xi^\sigma + T_{\sigma\nu}\nabla_\mu \xi^\sigma. \tag{3.1}$$

This formula becomes particularly simple if applied to the metric because $\xi^\sigma \nabla_\sigma g_{\mu\nu} = 0$ when using the Levi-Civita connection. After lowering the indices, one obtains

$$\mathcal{L}_\xi g_{\mu\nu} = \nabla_\mu \xi_\nu + \nabla_\nu \xi_\mu. \tag{3.2}$$

To mathematically define a symmetry, we need the concept of a family of integral curves[1], also called a **flow**. A flow can then be identified with a tangent vector field $\boldsymbol{\xi}$ which points along the tangent vector to the flow at each point $p \in \mathcal{M}$

$$\xi^\mu = \frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda}. \tag{3.3}$$

This flow is said to be an **isometry** if the metric looks the same at each point along a given flow line. Mathematically, this means that an isometry satisfies

$$\mathcal{L}_{\boldsymbol{\xi}}\, \boldsymbol{g} = 0, \tag{3.4}$$

which, according to (3.2) can be written

$$\nabla_\mu \xi_\nu + \nabla_\nu \xi_\mu = 0. \tag{3.5}$$

[1]The concept of integral curves was defined in Section 1.3.

This is the **Killing equation** and any tangent vector field $\boldsymbol{\xi}$ satisfying this equation is known as a **Killing vector**.

In practice, it is not always easy to find Killing vectors except in some peculiar situations. Indeed, let us explicitly write that the Lie derivative along some tangent vector $\boldsymbol{\xi}$ og the metric vanishes

$$\xi^\alpha \partial_\alpha g_{\mu\nu} + g_{\mu\alpha}\partial_\nu \xi^\alpha + g_{\alpha\nu}\partial_\mu \xi^\alpha = 0. \tag{3.6}$$

If we now try to solve this equation with a constant tangent vector $\boldsymbol{\xi}$ and assuming that the metric does not depend on a coordinate x^n *i.e.* $\frac{\partial g_{\mu\nu}}{\partial x^n} = 0$, then the above equation boils down to

$$\xi^\alpha \partial_\alpha g_{\mu\nu} = 0. \tag{3.7}$$

We then see that the tangent vector $\boldsymbol{\xi} = (0, ..., 1, ..., 0)$ where the 1 is at the n^{th} position is a Killing vector.

Moreover, it is easy to show that if $\boldsymbol{\xi}$ and $\boldsymbol{\chi}$ are Killing vectors, then any linear combination with non-varying constants $a\boldsymbol{\xi} + b\boldsymbol{\chi}$ and $[\boldsymbol{\xi}, \boldsymbol{\chi}]$ are also Killing vectors.

Example. • For a two dimensional sphere described in (θ, φ) coordinates, the metric reads[2]

$$\mathrm{d}s^2 = \mathrm{d}\theta^2 + \sin^2\theta\, \mathrm{d}\varphi^2. \tag{3.8}$$

The metric does not explicitly depend on φ so $\boldsymbol{\xi} = \partial_\varphi = (0, 1)$ is a Killing vector.

• For the Schwarzschild metris written in the usual spherical coordinates (t, r, θ, φ)

$$\mathrm{d}s^2 = -\left(1 - \frac{2M}{r}\right)\mathrm{d}t^2 + \left(1 - \frac{2M}{r}\right)^{-1}\mathrm{d}r^2 + r^2\mathrm{d}\Omega^2, \tag{3.9}$$

one sees that it does not depend on t nor φ so that $\boldsymbol{\xi} = \partial_t = (1, 0, 0, 0)$ and $\boldsymbol{\chi} = \partial_\varphi = (0, 0, 0, 1)$ are Killing vectors.

A Killing vector is a tangent vector $\boldsymbol{\xi}$ that satisfies the Killing equation

$$\mathcal{L}_{\boldsymbol{\xi}}\boldsymbol{g} = 0 \quad i.e. \quad \nabla_\mu \xi_\nu + \nabla_\nu \xi_\mu = 0. \tag{3.10}$$

If the metric does not dependent on a coordinate x^n, then

$$\xi^\mu = (0, ..., 0, \underbrace{1}_{n^{\text{th}}\ \text{position}}, 0, ..., 0), \tag{3.11}$$

is a Killing vector.

3.1.2 Conserved Charges

We are used to the fact that symmetries lead to conserved quantities. In the present context, the concept of "symmetries" is replaced by "symmetries of the metric", and we therefore expect conserved charges associated with the presence of Killing vectors. Here

[2] We set the radius $R = 1$.

are two important examples.

Let ξ^μ be a Killing vector and $x^\mu(\lambda)$ be a geodesic associated to the tangent vector X^μ, then the quantity

$$\xi_\mu X^\mu, \tag{3.12}$$

is a conserved quantity along the geodesic. Indeed,

$$\begin{aligned} X^\nu\nabla_\nu(\xi_\mu X^\mu) &= X^\nu(X^\mu\nabla_\nu\xi_\mu + \xi_\mu\nabla_\nu X^\mu) \\ &= \underbrace{X^\mu X^\nu}_{\text{symmetric}} \underbrace{\nabla_\nu\xi_\mu}_{\text{anti-symmetric}} + \xi_\mu \underbrace{X^\nu\nabla_\nu X^\mu}_{=0 \text{ geodesic}} \\ &= 0. \end{aligned} \tag{3.13}$$

Let ξ^μ be a Killing vector and $T^{\mu\nu}$ the covariantly conserved symmetric energy-momentum tensor, $\nabla_\mu T^{\mu\nu} = 0$, then the current

$$\xi_\nu T^{\mu\nu}, \tag{3.14}$$

is covariantly conserved. Indeed,

$$\nabla_\mu(\xi_\nu T^{\mu\nu}) = \underbrace{T^{\mu\nu}}_{\text{symmetric}} \underbrace{\nabla_\mu\xi_\nu}_{\text{anti-symmetric}} + \xi_\nu \underbrace{\nabla_\mu T^{\mu\nu}}_{=0} = 0. \tag{3.15}$$

Hence, as we now have a conserved current, we can associate with it a conserved charge in the way discussed above. The argument evidently relies on the fact that $T^{\mu\nu}$ is symmetric and covariantly conserved *i.e.* divergent-less.

Example. As an example of application, let us derive the energy conservation equation for the Schwarzschild metric where we assume[3] $\theta = \pi/2$

$$\mathrm{d}s^2 = -\left(1 - \frac{2M}{r}\right)\mathrm{d}t^2 + \left(1 - \frac{2M}{r}\right)^{-1}\mathrm{d}r^2 + r^2\mathrm{d}\varphi^2. \tag{3.16}$$

We know from te previous section that $\xi_t^\mu = \partial_t$ and $\xi_\varphi^\mu = \partial_\varphi$ are Killing vectors. Let us consider a tangent vector field[4] $\boldsymbol{U}$ such that $U_\mu U^\mu = \epsilon$ where $\epsilon = 0, \pm 1$. We then know that $-E = \xi_\mu^t U^\mu$ and $L = \xi_\mu^\varphi U^\mu$ are conserved quantities[5]. Denoting $U^\mu = (U^t, U^r, 0, U^\varphi)$, one obtains

$$\begin{aligned} g_{\mu\nu}\xi_t^\mu U^\nu &= g_{tt}U^t \\ g_{\mu\nu}\xi_\varphi^\mu U^\nu &= g_{\varphi\varphi}U^\varphi, \end{aligned} \tag{3.17}$$

which leads to

$$U^t = \frac{E}{1 - \frac{2M}{r}} \quad \text{and} \quad U^\varphi = \frac{L}{r^2}. \tag{3.18}$$

Writing $U_\alpha U^\alpha = g_{tt}(U^t)^2 + g_{rr}(U^r)^2 + g_{\varphi\varphi}(U^\varphi)^2 = \epsilon$ leads to

$$(U^r)^2 + \left(\frac{L^2}{r^2} - \epsilon\right)\left(1 - \frac{2M}{r}\right) = E^2. \tag{3.19}$$

[3] We assume planar trajectories.
[4] Namely the velocity of an object/particle.
[5] The total energy and the angular momentum respectively measured by observers at infinity.

Let ξ^μ be a Killing vector and $x^\mu(\lambda)$ be a geodesic associated to the tangent vector X^μ *i.e.* $X^\nu\nabla_\nu X^\mu = 0$, then the charge

$$\xi_\mu X^\mu, \tag{3.20}$$

is a conserved quantity along the geodesic.

Let ξ^μ be a Killing vector and $T^{\mu\nu}$ a covariantly conserved symmetric tensor[a], $\nabla_\mu T^{\mu\nu} = 0$, then the current

$$\xi_\nu T^{\mu\nu}, \tag{3.21}$$

is covariantly conserved *i.e.* divergent-less.

[a]For example the energy-momentum tensor.

3.1.3 Useful Identity Relating Curvature and Killing Vectors

In Riemannian geometry the rich interplay between symmetries and geometry is reflected in relations between the curvature tensor and Killing vectors of a metric.

Let ξ^μ be a Killing vector, then it must obey

$$\nabla_\mu\nabla_\nu\xi_\rho = R^\sigma_{\mu\nu\rho}\xi_\sigma, \tag{3.22}$$

where $R^\sigma_{\mu\nu\rho}$ is the Riemann tensor. Indeed, when applying three times the Riemann tensor identity

$$\nabla_\mu\nabla_\nu\xi^\sigma - \nabla_\nu\nabla_\mu\xi^\sigma = R^\sigma_{\rho\mu\nu}\xi^\rho, \tag{3.23}$$

and the Killing equation

$$\nabla_\mu\xi_\nu + \nabla_\nu\xi_\mu = 0, \tag{3.24}$$

one obtains

$$\begin{aligned}
\nabla_\mu\nabla_\nu\xi_\rho &= \nabla_\nu\nabla_\mu\xi_\rho - R^\sigma_{\rho\mu\nu}\xi_\sigma \\
&= -\nabla_\nu\nabla_\rho\xi_\mu - R^\sigma_{\rho\mu\nu}\xi_\sigma \\
&= -\nabla_\rho\nabla_\nu\xi_\mu - R^\sigma_{\rho\mu\nu}\xi_\sigma + R^\sigma_{\mu\nu\rho}\xi_\sigma \\
&= \nabla_\rho\nabla_\mu\xi_\nu - R^\sigma_{\rho\mu\nu}\xi_\sigma + R^\sigma_{\mu\nu\rho}\xi_\sigma \\
&= \nabla_\mu\nabla_\rho\xi_\nu - R^\sigma_{\rho\mu\nu}\xi_\sigma + R^\sigma_{\mu\nu\rho}\xi_\sigma - R^\sigma_{\nu\rho\mu}\xi_\sigma \\
&= -\nabla_\mu\nabla_\nu\xi_\rho - R^\sigma_{\rho\mu\nu}\xi_\sigma + R^\sigma_{\mu\nu\rho}\xi_\sigma - R^\sigma_{\nu\rho\mu}\xi_\sigma
\end{aligned} \tag{3.25}$$

so that

$$\nabla_\mu\nabla_\nu\xi_\rho = R^\sigma_{\mu\nu\rho}\xi_\sigma, \tag{3.26}$$

where we used the Bianchi identity.

Let ξ^μ be a Killing vector, then it must obey

$$\nabla_\mu \nabla_\nu \xi_\rho = R^\sigma_{\mu\nu\rho} \xi_\sigma, \tag{3.27}$$

where $R^\sigma_{\mu\nu\rho}$ is the Riemann tensor.

3.1.4 Maximal Symmetry and Constant Curvature

In order to understand how to define and characterise maximally symmetric spaces, we will need to obtain some more information about how Killing vectors can be classified. Our starting point is, as in the previous section, the identity reproduced here with the explicit x-dependence included for present purposes

$$\nabla_\mu \nabla_\nu \xi_\rho(x) = R^\sigma_{\mu\nu\rho}(x) \xi_\sigma(x). \tag{3.28}$$

In particular, this shows that the second derivatives of the Killing vector at a point x_0 are again expressed in terms of the value of the Killing vector itself at that point. This means that, remarkably, a Killing vector field $\xi_\mu(x)$ is completely and uniquely determined everywhere by the values of $\xi_\mu(x_0)$ and $\nabla_\mu \xi_\nu(x_0)$ at a single point x_0. Since, in an D-dimensional spacetime there can be at most D linearly independent vectors $(\xi_\mu(x_0))$ at a point, and at most $D(D-1)/2$ independent anti-symmetric matrices $(\nabla_\mu \xi_\nu(x_0))$, we reach the conclusion that an D-dimensional spacetime can have at most

$$D + \frac{D(D-1)}{2} = \frac{D(D+1)}{2}, \tag{3.29}$$

independent Killing vectors. A spacetime with this maximal number of Killing vectors is called **maximally symmetric**.

Example. The D-dimensional Minkowski spacetime is maximally symmetric. Note that $D(D+1)/2$ for $D = 4$ agrees with the dimension of the Poincaré group[6], the group of transformations that leave the Minkowski metric invariant. We can also cite the de Sitter and the anti-de Sitter spacetimes being maximally symmetric.

One can also show that maximally symmetric spacetimes have constant curvature $R =$ constant. For such spacetimes, one can show that the Riemann tensor takes the following form

$$R_{\mu\nu\rho\sigma} = \frac{R}{D(D-1)}(g_{\mu\rho}g_{\nu\sigma} - g_{\mu\sigma}g_{\nu\rho}). \tag{3.30}$$

[6] 3 rotations, 3 boosts, 3 translations and parity.

A D-dimensional spacetime is maximally symmetric is it has

$$\frac{D(D+1)}{2}, \tag{3.31}$$

independent Killing vectors. Those spacetimes have constant curvature $R =$ constant and the Riemann tensor can be written

$$R_{\mu\nu\rho\sigma} = \frac{R}{D(D-1)}(g_{\mu\rho}g_{\nu\sigma} - g_{\mu\sigma}g_{\nu\rho}). \tag{3.32}$$

3.2 Variational Calculus

Although we have met variational calculus when deriving the geodesic equation from the Euler-Lagrange equation, we will now introduce variational calculus of tensors. Ultimately, the aim is to be familiar with how to derive the Einstein equation starting from the Einstein-Hilbert action. This way of deriving the Einstein equation is more suitable for beyond classical General Relativity extensions because all our fundamental theories of physics are described by action principles.

3.2.1 Covariant Volume Element

In this section we will address the issue of generally covariant integration in a spacetime equipped with a metric. It is immediately apparent that the integral of a scalar $f(x)$ over spacetime $\int \mathrm{d}^D x\, f(x)$ is not generally covariant because one has to introduce the Jacobian of the coordinate transformation which in general is not equals to one.

Let us recall the standard tensorial transformation behaviour of the metric under coordinate transformations $x \to x'$,

$$g'_{\alpha\beta}(x') = \frac{\partial x^\mu}{\partial x'^\alpha}\frac{\partial x^\nu}{\partial x'^\beta} g_{\mu\nu}(x). \tag{3.33}$$

it follows that the absolute value of the metric determinant $|g| = |\det \boldsymbol{g}|$ does not transform like a scalar but instead transforms as

$$g' = \left|\det\left(\frac{\partial x}{\partial x'}\right)^2\right| g = \left|\det\left(\frac{\partial x'}{\partial x}\right)\right|^{-2} g. \tag{3.34}$$

In particular its square root $\sqrt{g}$ transforms as

$$\sqrt{g'} = \left|\det\left(\frac{\partial x'}{\partial x}\right)\right|^{-1} \sqrt{g}. \tag{3.35}$$

Therfore the combined expression $\mathrm{d}^D x\,\sqrt{g}$ is invariant under general coordinate transformation

$$\mathrm{d}^D x'\,\sqrt{g'} = \mathrm{d}^D x\,\sqrt{g}, \tag{3.36}$$

and can be used to define integrals of scalars $f(x)$ in a generally covariant way

$$\int \mathrm{d}^D x' \sqrt{g'}\, f(x') = \int \mathrm{d}^D x \sqrt{g}\, f(x). \tag{3.37}$$

Note that this is also frequently the quickest way to determine the volume element in non- Cartesian coordinates in Euclidean space.

Example. Let us derive the volume element of the three dimensional euclidean space in spherical coordinates (r, θ, φ) denoted by $\{x'^\mu\}$. The usual Cartesian coordinates are denoted $\{x^\mu\}$. Instead of laboriously determining the Jacobi matrix for the coordinate transformation, and then calculating its determinant, all one needs to know is the metric

$$\mathrm{d}s^2 = \mathrm{d}r^2 + r^2 \mathrm{d}\theta^2 + r^2 \sin^2\theta \mathrm{d}\varphi^2. \tag{3.38}$$

Computing the determinant of the metric is straightforward because it is diagonal $g = r^4 \sin^2\theta$ and therefore we have

$$\mathrm{d}^3 x = \sqrt{g}\, \mathrm{d}^3 x' = r^2 \sin\theta\, \mathrm{d}r\, \mathrm{d}\theta\, \mathrm{d}\varphi, \tag{3.39}$$

which is of course the standard result.

The covariant volume element for integration is $\mathrm{d}^D x \sqrt{g}$ which is invariant under general coordinate transformation $x \to x'$

$$\mathrm{d}^D x' \sqrt{g'} = \mathrm{d}^D x \sqrt{g}, \tag{3.40}$$

and can be used to define integrals of scalars $f(x)$ in a generally covariant way

$$\int \mathrm{d}^D x' \sqrt{g'}\, f(x') = \int \mathrm{d}^D x \sqrt{g}\, f(x). \tag{3.41}$$

3.2.2 Divergence Theorem

Gauss's theorem, also known as the **divergence theorem**, states that if you integrate a total derivative, you get a boundary term. There is a particular version of this theorem in curved space. Before dive in, let us first state the following result.

Lemma. The contraction of the Christoffel symbols can be written as

$$\Gamma^\mu_{\mu\nu} = \frac{1}{\sqrt{g}} \partial_\nu \sqrt{g}, \tag{3.42}$$

where $g = \det \boldsymbol{g}$. Note that for Lorentzian manifolds, the metric determinant is negative and so one needs to perform the following replacement $\det \boldsymbol{g} \to \det |\boldsymbol{g}|$. Indeed, from the explicit relation between the Christoffel symbols and the metric, one obtains

$$\Gamma^\mu_{\mu\nu} = \frac{1}{2} g^{\mu\rho} \partial_\nu g_{\mu\rho} = \frac{1}{2} \mathrm{Tr}\left[\boldsymbol{g}^{-1} \partial_\nu \boldsymbol{g}\right] = \frac{1}{2} \mathrm{Tr}\left[\partial_\nu \log \boldsymbol{g}\right]. \tag{3.43}$$

However, there is a useful identity for the log of any diagonalisable matrix $\boldsymbol{A}$. They obey

$$\mathrm{Tr}\left[\log \boldsymbol{A}\right] = \log \det \boldsymbol{A}. \tag{3.44}$$

This is clearly true for a diagonal matrix, since the determinant is the product of eigenvalues while the trace is the sum. But both trace and determinant are invariant under conjugation, so this is also true for diagonalisable matrices. Applying it to our metric formula above, we have

$$\Gamma^{\mu}_{\mu\nu} = \frac{1}{2}\mathrm{Tr}\left[\partial_\nu \log \boldsymbol{g}\right] = \frac{1}{2}\partial_\nu \log \det \boldsymbol{g} = \frac{1}{2}\frac{1}{\det \boldsymbol{g}}\partial_\nu \det \boldsymbol{g} = \frac{1}{\sqrt{\det \boldsymbol{g}}}\partial_\nu \sqrt{\det \boldsymbol{g}}, \tag{3.45}$$

which is the claimed result. With this at hand, we can now prove the following theorem.

Theorem. Consider a region of a D-dimensional manifold $\mathcal{M}$ with boundary $\partial\mathcal{M}$. Let n^μ be an outward-pointing, unit vector orthogonal to $\partial\mathcal{M}$. Then, for any tangent vector field X^μ on $\mathcal{M}$, we have

$$\int_{\mathcal{M}} \mathrm{d}^D x \sqrt{g}\, \nabla_\mu X^\mu = \int_{\partial\mathcal{M}} \mathrm{d}^{D-1} x \sqrt{\gamma}\, n_\mu X^\mu, \tag{3.46}$$

where γ_{ij} is the pull-back of the metric to $\partial\mathcal{M}$ and $\gamma = \det \boldsymbol{\gamma}$.

Proof. Using the lemma above, the integrand is

$$\sqrt{g}\, \nabla_\mu X^\mu = \sqrt{g}(\partial_\mu X^\mu + \Gamma^\mu_{\mu\nu} X^\nu) = \sqrt{g}(\partial_\mu X^\mu + X^\nu \frac{1}{\sqrt{g}}\partial_\nu \sqrt{g}) = \partial_\nu(\sqrt{g}\, X^\mu), \tag{3.47}$$

so that the integral is

$$\int_{\mathcal{M}} \mathrm{d}^D x \sqrt{g}\, \nabla_\mu X^\mu = \int_{\mathcal{M}} \mathrm{d}^D x\, \partial_\nu(\sqrt{g}\, X^\mu), \tag{3.48}$$

which is now an integral of an ordinary partial derivative so we can apply the usual divergence, also called Stokes' theorem, that we are familiar with. It remains only to evaluate what is happening at the boundary $\partial\mathcal{M}$. For this, it is useful to pick coordinates so that the boundary $\partial\mathcal{M}$ is a surface of constant x^n. Furthermore, we will restrict to metrics of the form[7]

$$g_{\mu\nu} = \begin{pmatrix} \gamma_{ij} & 0 \\ 0 & 1 \end{pmatrix}. \tag{3.49}$$

Then by our usual result of integration, we have

$$\int_{\mathcal{M}} \mathrm{d}^D x\, \partial_\nu(\sqrt{g}\, X^\mu) = \int_{\partial\mathcal{M}} \mathrm{d}^{D-1} x\, \sqrt{\gamma}\, X^n. \tag{3.50}$$

The unit normal vector n^μ is given by $n^\mu = (0, 0, ..., 1)$, which satisfies $g_{\mu\nu} n^\mu n^\nu = 1$ as it should. We then have $n_\mu = g_{\mu\nu} n^\nu = (0, 0, ..., 1)$ so wa can write

$$\int_{\mathcal{M}} \mathrm{d}^D x\, \partial_\nu(\sqrt{g}\, X^\mu) = \int_{\partial\mathcal{M}} \mathrm{d}^{D-1} x\, \sqrt{\gamma}\, n_\mu X^\mu, \tag{3.51}$$

which is the result we need. As the final expression is a covariant quantity, it is true in general.

[7]We construct a metric γ_{ij} for the boundary hypersurface.

The contraction of the Christoffel symbols can be written

$$\Gamma^{\mu}_{\mu\nu} = \frac{1}{\sqrt{g}} \partial_\nu \sqrt{g}, \tag{3.52}$$

where[a] $g = \det \boldsymbol{g}$.

Consider a region of a D-dimensional manifold $\mathcal{M}$ with boundary $\partial\mathcal{M}$. Let n^μ be an outward-pointing, unit vector orthogonal to $\partial\mathcal{M}$. Then, for any tangent vector field X^μ on $\mathcal{M}$, we have

$$\int_{\mathcal{M}} \mathrm{d}^D x \sqrt{g}\, \nabla_\mu X^\mu = \int_{\partial\mathcal{M}} \mathrm{d}^{D-1} x \sqrt{\gamma}\, n_\mu X^\mu, \tag{3.53}$$

where γ_{ij} is the pull-back of the metric to $\partial\mathcal{M}$ and $\gamma = \det \boldsymbol{\gamma}$.

[a] Note that for Lorentzian manifolds, the metric determinant is negative and so one needs to perform the following replacement $\det \boldsymbol{g} \to \det |\boldsymbol{g}|$.

3.2.3 Einstein-Hilbert Action

Gravity is also a fundamental theory that can be described by action principles. The straight-jacket of differential geometry places enormous restrictions on the kind of actions that we can write down. These restrictions ensure that the action is something intrinsic to the metric itself, rather than depending on our choice of coordinate.

In General Relativity, the gravitational field is identified with a metric $g_{\mu\nu}$ on a four dimensional Lorentzian manifold that we call spacetime. To build the **Einstein-Hilbert action** that governs the metric dynamics, we know, from a previous section, that we need the covariant volume element to integrate over a manifold. Furthermore, given that we only have the metric to play with, the simplest scalar function is the Ricci scalar R. This motivates us to consider the wonderful concise action

$$S = \int \mathrm{d}^4 x \sqrt{-g}\, R, \tag{3.54}$$

which is the famous Einstein-Hilbert action[8]. Note that the minus sign under the square root arises because we are in a Lorentzian spacetime. As a quick sanity check, recall that the Ricci tensor takes the schematic form $R \sim \partial\Gamma + \Gamma\Gamma$ while the Levi-Civita connection itself is $\Gamma \sim \partial g$. This means that the Einstein-Hilbert action is second order in derivatives, just like other actions we consider in physics.

Note that written in this way, the Einstein-Hilbert action is suitable to consider that classical General Relativity is just an approximated theory of a more general theory with the following action

$$S = \int \mathrm{d}^4 x \sqrt{-g}\, (R + \sigma_2 R^2 + \sigma_3 R^3 + \dots), \tag{3.55}$$

[8] Note that a dimensional analysis requires that the physical action must be multiplied by $c^3/(16\pi G)$.

where σ_i are constants. For example truncating the series expansion up to second order *i.e.* keeping only the $\sim R^2$ term gives rise to the so-called Starobinsky potential which can describe inflation during an early stage of the Universe[9].

In General Relativity, the gravitational field is identified with a metric $g_{\mu\nu}$ on a four dimensional Lorentzian manifold that we call spacetime. The metric dynamics is governed by the Einstein-Hilbert action

$$S = \int \mathrm{d}^4x \sqrt{-g}\, R. \tag{3.56}$$

3.2.4 Action Variation

In this section we would like to determine the Euler-Lagrange equation arising from the Einstein-Hilbert action. We do this in the usual way, by starting with some fixed metric $g_{\mu\nu}(x)$ and seeing how the action changes when we shift $g_{\mu\nu}(x) \to g_{\mu\nu}(x) + \delta g_{\mu\nu}(x)$. We then try to write the integrand of δS as proportional to $\delta g^{\mu\nu}$.

Writing the Ricci scalar as $R = g^{\mu\nu}R_{\mu\nu}$, the Einstein-Hilbert action changes as

$$\delta S = \int \mathrm{d}^4x \left([\delta\sqrt{-g}]g^{\mu\nu}R_{\mu\nu} + \sqrt{-g}[\delta g^{\mu\nu}]R_{\mu\nu} + \sqrt{-g}\, g^{\mu\nu}\,[\delta R_{\mu\nu}]\right). \tag{3.57}$$

The aim is then to derive the variation of the inverse metric, the covariant volume element and the Ricci tensor.

It turns out that it is slightly easier to think of the variation in terms of the inverse metric $\delta g^{\mu\nu}$. This is equivalent to the variation of the metric $\delta g_{\mu\nu}$, the two are related by $g_{\rho\mu}g^{\mu\nu} = \delta^\nu_\rho$ so that

$$(\delta g_{\rho\mu})g^{\mu\nu} + g_{\rho\mu}\delta g^{\mu\nu} = 0 \quad i.e. \quad \delta g^{\mu\nu} = -g^{\mu\rho}g^{\nu\sigma}\delta g_{\rho\sigma}. \tag{3.58}$$

The middle term in (3.57) is already proportional to $\delta g^{\mu\nu}$. We now deal with the first and third terms in turn.

First, we use the standard trick $\log \det \boldsymbol{A} = \mathrm{Tr}[\log \boldsymbol{A}]$ for any diagonalisable matrix $\boldsymbol{A}$ to write

$$\frac{1}{\det \boldsymbol{A}}\,\delta(\det \boldsymbol{A}) = \mathrm{Tr}[\boldsymbol{A}^{-1}\delta \boldsymbol{A}]. \tag{3.59}$$

Applying this result to the metric, we have

$$\delta\sqrt{-g} = \frac{1}{2}\frac{1}{\sqrt{-g}}(-g)g^{\mu\nu}\delta g_{\mu\nu} = \frac{1}{2}\sqrt{-g}g^{\mu\nu}\delta g_{\mu\nu}. \tag{3.60}$$

Using $g^{\mu\nu}\delta g_{\mu\nu} = -g_{\mu\nu}\delta g^{\mu\nu}$, one obtains the variation of the covariant volume element

[9]However, this potential has been proved wrong by the latest Planck satellite data of the Cosmic Microwave Background.

$$\delta\sqrt{-g} = -\frac{1}{2}\sqrt{-g}\, g_{\mu\nu}\delta g^{\mu\nu}. \tag{3.61}$$

We now claim that that the final term $g^{\mu\nu}\delta R_{\mu\nu}$ is a total derivative (boundary term) and can be written as[10]

$$g^{\mu\nu}\delta R_{\mu\nu} = \nabla_\mu X^\mu \quad \text{with} \quad X^\mu = g^{\rho\nu}\delta\Gamma^\mu_{\rho\nu} - g^{\mu\nu}\delta\Gamma^\rho_{\nu\rho}. \tag{3.62}$$

using all the previous results, the variation of the action can then be written

$$\delta S = \int \mathrm{d}^4x\,\sqrt{-g}\left[\left(R_{\mu\nu} - \frac{1}{2}R\, g_{\mu\nu}\right)\delta g^{\mu\nu} + \nabla_\mu X^\mu\right]. \tag{3.63}$$

This final term is a total derivative and by the divergence theorem, stated in a previous section, we ignore it. Requiring that the action is extremised $\delta S = 0$, one obtains the metric equation of motion[11]

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}R\, g_{\mu\nu} = 0. \tag{3.64}$$

Varying the Einstein-Hilbert action reads

$$\delta S = \int \mathrm{d}^4x\,\left([\delta\sqrt{-g}]g^{\mu\nu}R_{\mu\nu} + \sqrt{-g}[\delta g^{\mu\nu}]R_{\mu\nu} + \sqrt{-g}\, g^{\mu\nu}\,[\delta R_{\mu\nu}]\right). \tag{3.65}$$

Varying the covariant volume element and the Ricci tensor reads

$$\begin{aligned} \delta\sqrt{-g} &= -\frac{1}{2}\sqrt{-g}\, g_{\mu\nu}\delta g^{\mu\nu} \\ g^{\mu\nu}\delta R_{\mu\nu} &= \nabla_\mu X^\mu \quad \text{with} \quad X^\mu = g^{\rho\nu}\delta\Gamma^\mu_{\rho\nu} - g^{\mu\nu}\delta\Gamma^\rho_{\nu\rho}. \end{aligned} \tag{3.66}$$

The last variation turns out to be a boundary term. Using all the previous results, the variation of the action is written

$$\delta S = \int \mathrm{d}^4x\,\sqrt{-g}\left[\left(R_{\mu\nu} - \frac{1}{2}R\, g_{\mu\nu}\right)\delta g^{\mu\nu} + \nabla_\mu X^\mu\right]. \tag{3.67}$$

Ignoring the boundary term and requiring the action to be extremised $\delta S = 0$, one obtains the Einstein field equation in vacuum

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}R\, g_{\mu\nu} = 0. \tag{3.68}$$

Note that they simplify somewhat; if we contract the last equation with $g^{\mu\nu}$, we find that $R = 0$. Substituting this back in, the vacuum Einstein equation is simply the requirement that the metric is Ricci flat

$$R_{\mu\nu} = 0. \tag{3.69}$$

[10]We do not prove this result because it is very long. Note that one can easily derive this result in normal coordinates but these are not introduced in these lecture notes. For interested people, I found a Youtube video that explicitly makes this derivation.

[11]In vaccum.

Note that we happily discarded the boundary term, a standard practice whenever we invoke the variational principle. It turns out that there are some situations in General Relativity where we should not be quite so cavalier. In such circumstances, one can be more careful by invoking the so-called Gibbons-Hawking boundary term.

The Essential in a Scheme