MA4K5: Introduction to Mathematical Relativity

Dr. Claude Warnick

November 11, 2017

Abstract

One of the crowning achievements of modern physics is Einstein's theory of general relativity, which describes the gravitational field to a very high degree of accuracy. As well as being an astonishingly accurate physical theory, the study of general relativity is also a fascinating area of mathematical research, bringing together aspects of differential geometry and PDE theory. In this course, I will introduce the basic objects and concepts of general relativity without assuming a knowledge of special relativity. The ultimate goal of the course will be a discussion of the Cauchy problem for the vacuum Einstein equations, including a statement of the relevant well-posedness theorems and a discussion of their relevance. We will take a 'field theory' approach to the subject, emphasising the deep connection between Lorentzian geometry and hyperbolic PDE. In contrast to the course PX436 General Relativity offered by the department of physics, we concentrate on the mathematical structure of the theory rather than its physical implications.

By the end of the module, students should be able to:

- Understand how the Minkowski geometry and Lorentz group arise from considerations of signal propagation for the scalar wave equation.
- Understand the basics of Lorentzian geometry: the metric; causal classification of vectors; connection and curvature; hypersurface geometry; conformal compactifications; the d'Alembertian operator.
- Be able to state the well-posedness theorems for the Cauchy problem for the Einstein equations and sketch the proof of local well posedness.

Contents

Acknowledgements

Introduction

A brief history of light

Light moves fast. Famously fast. So fast, in fact, that it's extremely difficult to observe that its speed, c, is finite. For this reason many people through history have, understandably, concluded that the speed of light is infinite. Debate on the matter dates back at least to the ancient Greeks, and the finiteness of c was only really established towards the end of the 17^{th} century.

The first definitive experimental proof that light has a finite speed was given by Rømer in 1676. His argument was based on measurements of Jupiter's moon Io. At some times in the year, Earth is moving towards Jupiter, while at others it is moving away. Rømer observed what we would now call a Doppler shift in the orbital frequency of Io. When the Earth moves towards Jupiter, the frequency of its orbit appears to increase, while when Earth is moving away from Jupiter the frequency decreases. By combining his measurements of the frequency change with estimates of the diameter of Earth's orbit, Rømer argued for a value (in modern units)

$$c \simeq 2.2 \times 10^8 \; ms^{-1},$$

which is pretty good considering our modern value of $c = 3.00 \times 10^8 \; ms^{-1}$. Although controversial at first, further measurements confirmed Rømer's demonstration that c is indeed finite.

At the time, there were two competing theories regarding the nature of light. Isaac Newton favoured the 'corpuscular theory of light', according to which light consists of particles. The competing theory of Huygens instead described light as a wave, propagating through the 'luminiferous æther'. It was not until 1804 that Young undertook his famous 'twin slits' experiment and demonstrated the wave character of light.

The next step in understanding light came as part of one of the great achievements of 19^{th} century physics. This was the unification in the early 1860s, by Maxwell, of the electrical and magnetic forces. Maxwell's equations in a vacuum have the form:

$$\nabla \cdot \boldsymbol{E} = 0, \tag{1}$$

$$\frac{\partial \boldsymbol{B}}{\partial t} + \nabla \times \boldsymbol{E} = 0, \tag{2}$$

$$\nabla \cdot \boldsymbol{B} = 0, \tag{3}$$

$$\epsilon_0 \frac{\partial \boldsymbol{E}}{\partial t} - \frac{1}{\mu_0} \nabla \times \boldsymbol{B} = 0. \tag{4}$$

Here $\boldsymbol{E}, \boldsymbol{B}$ are the electric and magnetic fields and μ_0, ϵ_0 are two constants, known as the permeability and permittivity of free space. From these constants, we can form a combination, $(\epsilon_0 \mu_0)^{-\frac{1}{2}}$, which somewhat suggestively has dimensions of speed. One of Maxwell's great contributions was to show that the system of PDEs (1–4) admits propagating wave solutions, with speed $c = (\epsilon_0 \mu_0)^{-\frac{1}{2}}$, and to identify these electromagnetic waves with light.

Exercise(∗). Consider a system of particles of mass m_i at positions $\boldsymbol{r}_i(t) \in \mathbb{R}^3$. Suppose that the particle j exerts a force $\boldsymbol{F}_{ij}$ on the particle i, where $\boldsymbol{F}_{ij} = \boldsymbol{F}_{ij}(\boldsymbol{r}_i - \boldsymbol{r}_j)$ depends only on the relative separation of the particles. Show that Newton's equation of motion for the system:

$$m_i \frac{d^2 \boldsymbol{r}_i}{dt^2} = \sum_j \boldsymbol{F}_{ij}$$

is invariant under the Galilean boost $\boldsymbol{r}_i \to \boldsymbol{r}_i(t) + \boldsymbol{v}t$.

As a means of describing the nature of light, Maxwell's equations were a triumph. Their interpretation gave rise to certain puzzles, however. There is a definite speed, c, present in Maxwell's equations. What should this speed be measured relative to? It had been known since Galileo and Newton that the laws of mechanics do not define a definite frame of reference: two identical mechanical systems moving at a constant speed relative to one another cannot be distinguished. Not so for Maxwell's equations: changing to a different frame moving at a constant speed by the Galilean boost which leaves Newton's mechanics invariant does *not* preserve the Maxwell equations.

One possible resolution to this issue was to postulate the existence of the 'aether', some fluid-like substance through which light moves (and relative to which the speed of light is c). This however gives rise to more questions. If light propagates as waves in some aether, what are its properties? How does it interact with moving bodies? Is it like a fluid, flowing around solid objects and dragged along by them when they move? Or does it flow through matter without interacting with it?

Several experiments were made towards the end of the 19th century to try and answer these questions. The most well known of these was the Michelson-Morley experiment, which aimed to measure the speed of the earth relative to the aether. To do this, they measured the speed of light at various times of day and at different times of the year.

If light does behave like water waves on some background fluid, then one would expect to see directional and seasonal variation in the speed of light, with it apparently moving faster in the direction of the 'aether wind'. No such effect was observed by Michelson and

Morley. Several people helped resolve this paradox, with particular contributions due to Fitzgerald, Lorentz and Poincaré, who between them showed that it's possible to make a change of coordinates preserving the form of Maxwell's equations and representing a shift to a frame in uniform motion, but only if one transforms both time and space variables simultaneously. In 1905, Einstein was able to derive these transformations from the relativity postulate (that physics is the same in two frames which are in uniform relative motion) and the constancy of the speed of light. The resulting theory of space and time is known as Special Relativity.

Gravity and General Relativity

At the start of the 20th century, the prevailing theory of gravitation was that of Newton. In this theory, the gravitational field is represented by a function, Φ, which solves the Poisson equation:

$$\Delta\Phi = 4\pi G\rho, \tag{5}$$

with $\rho(x,t)$ the density of matter. The gravitational force on a particle of mass m is then give by:

$$\boldsymbol{F} = -m\boldsymbol{\nabla}\Phi. \tag{6}$$

This theory of gravitation successfully describes almost all of the gravitational phenomena that are observable in our solar system. However, it is not compatible with Einstein's theory of Special Relativity. The reason for this is that the field Φ exhibits an infinite speed of propagation: a change in ρ instantly causes Φ to change *everywhere in space simultaneously*. Coupling this type of theory to Special Relativity introduces many paradoxes.

Finding a way to reconcile his theory of Special Relativity with Newton's theory of gravity took Einstein 10 years. The crucial observation that permitted him to find a relativistic theory of gravitation is the *principle of equivalence*. All bodies in a gravitational field accelerate at the same rate *regardless of their mass*. Einstein realised that this phenomenon could be explained if freely falling bodies follow the geodesics of a curved geometry with metric g. This postulate replaces the equation (6).

The curved metric g encodes the gravitational field. Einstein's field equations:

$$Ric_g - \frac{1}{2}Rg = \frac{8\pi G}{c^2}T \tag{7}$$

relate the curvature of the geometry to the density of matter, encoded in T, and are the replacement for (5).

In this course, we shall study some of the mathematical underpinnings of the theories of special and general relativity, as they are now understood. We will start off by studying the linear wave equation. For our purposes, this is a slightly simplified version of Maxwell's equations. Through the study of the wave equation, we will be lead to a study of the Lorentzian geometry of Minkowski space. We will then introduce the concepts of curvature, which will allow us to formulate Einstein's equations. We will finally move on to discuss the solvability of Einstein's equations.

Chapter 1

The Wave Equation and Special Relativity

1.1 The wave equation from Maxwell's equations

In this course we are going to take a PDE based approach to relativity. We will begin with exploring special relativity in the context of the wave equation:

$$-\frac{\partial^2 u}{\partial t^2} + c^2 \Delta u = 0.$$

This provides a convenient proxy for the study of Maxwell's equations. We could instead study Maxwell's equations directly, but since these are a system of PDEs for 6 components of the electromagnetic field they can be a bit unwieldy.

In fact, Maxwell's equations are closely related to the wave equation, as our first result will establish:

Lemma 1.1. *Let us denote* $S_T := \mathbb{R}^3 \times (-T, T)$. *Suppose* $\boldsymbol{E}, \boldsymbol{B} \in C^2(S_T)$ *satisfy Maxwell's equations. Then each component of* $\boldsymbol{E}, \boldsymbol{B}$ *satisfies the wave equation:*

$$-\frac{\partial^2 E_i}{\partial t^2} + c^2 \Delta E_i = 0$$
$$-\frac{\partial^2 B_i}{\partial t^2} + c^2 \Delta B_i = 0$$

in S_T, *where* $c = (\mu_0 \epsilon_0)^{-\frac{1}{2}}$.

Proof. We start with (2) and differentiate in time to find

$$-\frac{\partial^2 \boldsymbol{B}}{\partial t^2} - \boldsymbol{\nabla} \times \frac{\partial \boldsymbol{E}}{\partial t} = 0.$$

Using (4) to replace the term involving $\dot{\boldsymbol{E}}$, we have

$$-\frac{\partial^2 \boldsymbol{B}}{\partial t^2} - \frac{1}{\mu_0 \epsilon_0} \boldsymbol{\nabla} \times (\boldsymbol{\nabla} \times \boldsymbol{B}) = 0.$$

Now, a standard vector calculus identity tells us that $\boldsymbol{\nabla} \times (\boldsymbol{\nabla} \times \boldsymbol{A}) = \boldsymbol{\nabla} (\boldsymbol{\nabla} \cdot \boldsymbol{A}) - \Delta \boldsymbol{A}$ for any C^2 vector field $\boldsymbol{A}$. Making use of this, together with (3), we deduce the result. □

Exercise(∗). Complete the proof by showing that the components of $\boldsymbol{E}$ obey the wave equation.

Later on in the course we shall be able to show a converse to this result, namely that we can find a solution of Maxwell's equations by solving the wave equation for each component.

At this stage, it's useful to assume that we are using units in which $c = 1$. For example, we can take the second as our unit of time and the light-second as our unit of length. This is convenient as it saves us carrying around a constant in our formulae. If you want to replace c it's always possible to do so by thinking about what units various quantities ought to carry.

1.2 The Cauchy problem for the wave equation

For the rest of this chapter, we shall be discussing solutions of the wave equation. We will start by considering the Cauchy problem. In general, the Cauchy problem for a PDE consists of specifying some data on a given surface and then trying to find a unique solution of the PDE in a neighbourhood of that surface. We will discuss this much more thoroughly when we come to the Cauchy-Kovalevskaya theorem. For now, let us consider what data we might expect to have to specify on the surface $\Sigma_0 := \mathbb{R}^3 \times \{0\}$ in order to find a unique solution of the wave equation:

$$-\frac{\partial^2 u}{\partial t^2} + \Delta u = 0 \tag{1.1}$$

in $S_T := \mathbb{R}^3 \times (-T, T)$. In order to see what data might be necessary to solve this problem, we can try and write u as a formal[1] power series in t about $t = 0$. That is, to try and find coefficients u_n so that the formal power series:

$$u(x,t) = \sum_{n=0}^{\infty} u_n(x) \frac{t^n}{n!}$$

solves (1.1). Putting this series into the equation, we find that we can cancel the terms order by order if:

$$-u_{n+2}(x) + \Delta u_n(x) = 0, \qquad n = 0, 1, \ldots$$

In other words, once we have specified $u_0(x) = u(x,0)$ and $u_1(x) = u_t(x,0)$, at least in principle we can find the rest of the terms in the formal power series by repeated differentiation. In other words, we expect that the correct *Cauchy data* for the wave equation on Σ_0 are the values of $u|_{\Sigma_0}$ and $u_t|_{\Sigma_0}$.

The formal power series argument above, while suggestive, is not terribly useful. For the solutions we shall ultimately be interested in, the series expansion above will not converge, so any arguments based on manipulating these series are rather suspect. We need a better tool before we can understand solutions of the wave equation. In particular, we require some *a priori* estimates for the solutions. Recall from the Theory of PDE

[1] in this context, 'formal' means that we will ignore issues of convergence.

course that an a priori estimate is a an estimate that we can deduce directly from the equation, *without having to write down a solution*, i.e. they are estimates that must hold for all solutions. In the Theory of PDE course, the main a priori estimates we used were the Maximum Principle for Laplace's equation and for the heat equation. For the wave equation *no maximum principle exists*, instead we make use of energy estimates.

Theorem 1.1 (Basic energy estimate for the wave equation)**.** *Let* $S_T := \mathbb{R}^3 \times (-T, T)$ *and* $\Sigma_t = \mathbb{R}^3 \times \{t\}$. *Suppose that* $u \in C^2(S_T)$ *solves (1.1) in* S_T*, and that there exists* R *such that* $u(x,t) = 0$ *for* $|x| > R$. *Then if we define*

$$E[u](t) := \frac{1}{2}\int_{\Sigma_t} \left(u_t^2 + |\nabla u|^2\right) d\sigma = \frac{1}{2}\int_{\mathbb{R}^3} \left(u_t(x,t)^2 + |\nabla u(x,t)|^2\right) dx,$$

we have

$$\frac{dE[u]}{dt} = 0.$$

Proof. The assumptions on $u(x,t)$ imply that we can restrict the range of the integration in E to $B_{2R}(0)$ and that that we can differentiate $E(t)$ with respect to time and pass the time derivative under the integral[2] to obtain

$$\begin{aligned}\frac{dE[u]}{dt} &= \int_{B_{2R}(0)} (u_t u_{tt} + \nabla u \cdot \nabla u_t)\, dx \\ &= \int_{B_{2R}(0)} (u_t u_{tt} - \Delta u u_t)\, dx + \int_{\partial B_{2R}(0)} u_t \nabla u \cdot \boldsymbol{n}\, d\sigma\end{aligned}$$

Here we have used the vector calculus identity $\nabla \cdot (f\nabla g) = \nabla f \cdot \nabla g + f\Delta g$ and then applied the divergence theorem. The first integral vanishes, because the integrand is proportional to $u_{tt} - \Delta u$ which is zero since u obeys the wave equation. The second integral vanishes because $u = 0$ for $|x| > R$. The result follows. □

Later on, we shall show that the assumption that solutions vanish outside a sufficiently large ball is in fact a reasonable one, because solutions to the wave equation exhibit *finite speed of propagation.* In the context of relativity, this is an analogue of the statement that no signal may travel faster than the speed of light.

From this result, we can immediately deduce that a solution of the wave equation is indeed uniquely determined by our proposed Cauchy data, namely $u|_{\Sigma_0}$ and $u_t|_{\Sigma_0}$.

Corollary 1.2. *Suppose* $u, v \in C^2(S_T)$ *solve (1.1) in* S_T*, and that there exists* R *such that* $u(x,t) = v(x,t) = 0$ *for* $|x| > R$. *Suppose further that:*

$$u|_{\Sigma_0} = v|_{\Sigma_0}, \qquad u_t|_{\Sigma_0} = v_t|_{\Sigma_0}.$$

Then $u = v$ *in* S_T.

[2] Check that you understand why this is true. You may wish to look at the appendix of the notes for the course MA3G1.

Proof. Consider $w = u - v$. This solves (1.1) in S_T and is in $C^2(S_T)$. Moreover, $w|_{\Sigma_0} = w_t|_{\Sigma_0} = 0$. Thus $E[w](0) = 0$. By Theorem 1.1 we have that $E[w](t) = 0$ for all $-T < t < T$. As a consequence, we have $|\nabla w| = |w_t| = 0$ on S_T, which together with the fact that w vanishes for large x implies that $w = 0$ on S_T. □

Our energy estimate has shown that specifying u and u_t on a surface of constant time is enough to ensure the solution to the Cauchy problem, if it exists, is unique. In fact, the energy estimate gives us a statement of continuity: the L^2-norms of $u_t(\cdot, t)$ and $|\nabla u|(\cdot, t)$ are controlled by the initial data. More precisely, we say that two solutions u, v as in Corollary 1.3 are $\epsilon-$close in the energy norm at time t if:

$$E[u - v](t) \leq \epsilon.$$

We clearly have

Corollary 1.3. *Suppose $u, v \in C^2(S_T)$ solve (1.1) in S_T, and that there exists R such that $u(x, t) = v(x, t) = 0$ for $|x| > R$. Suppose further that u and v are initially $\epsilon-$close in the energy norm. Then they remain $\epsilon-$close for all times $t \in (-T, T)$.*

We can improve the control over the solution to control higher derivatives:

Exercise 1.1. Let $S_T := \mathbb{R}^3 \times (-T, T)$ and $\Sigma_t = \mathbb{R}^3 \times \{t\}$. Fix $k \in \mathbb{N}$. Suppose that $u \in C^{2+k}(S_T)$, solves the wave equation (1.1) in S_T, and that there exists R such that $u(x, t) = 0$ for $|x| > R$. Define $u_0 := u|_{\Sigma_0}$ and $u_1 := u_t|_{\Sigma_0}$.

a) By deriving an equation for $\nabla_i u$ for $i = 1, 2, 3$ show that[3]

$$\frac{1}{2}\int_{\Sigma_t}\left(|\nabla u_t|^2 + \left|\nabla^2 u\right|^2\right) d\sigma = \frac{1}{2}\int_{\mathbb{R}^3}\left(|\nabla u_1|^2 + \left|\nabla^2 u_0\right|^2\right) dx$$

for $-T < t < T$.

b) Deduce that:

$$\frac{1}{2}\int_{\Sigma_t}\left(\left|\nabla^k u_t\right|^2 + \left|\nabla^{k+1} u\right|^2\right) d\sigma = \frac{1}{2}\int_{\mathbb{R}^3}\left(\left|\nabla^k u_1\right|^2 + \left|\nabla^{k+1} u_0\right|^2\right) dx.$$

for $-T < t < T$.

We have thus established that a solution, if it exists, to the Cauchy problem for the wave equation is unique and moreover that the solution depends continuously (in an appropriate sense) on the initial data. The final aspect of well posedness that remains to prove is that a solution does in fact exist. Rather than prove this now, we shall postpone our discussion to a later date, and just state a well posedness result.

[3] Here $\left|\nabla^2 u\right|^2 = \sum_{i,j} \nabla_i \nabla_j u \nabla_i \nabla_j u$, etc.

Theorem 1.4 (Well posedness of the wave equation)**.** *Suppose that* $u_0, u_1 \in C_c^\infty(\mathbb{R}^3)$*. Then there exists a unique solution* $u \in C^\infty(\mathbb{R}^3 \times \mathbb{R})$ *to the Cauchy problem:*

$$\begin{cases} -u_{tt} + \Delta u = 0 & \text{in } \mathbb{R}^3 \times \mathbb{R}, \\ u = u_0, \quad u_t = u_1 & \text{on } \mathbb{R}^3 \times \{0\}. \end{cases}$$

Moreover, if supp $u_i \subset B_R(0)$*, for* $i = 1, 2$*, then supp* $u(\cdot, t) \subset B_{R+|t|}(0)$*.*

Remarks: we have stated the theorem for smooth functions. Finite regularity versions of this theorem can be stated, however, if we work with spaces of C^k functions, then we have to assume more regularity on the initial data than we are able to recover for the solution. We say that the C^k-norms are not propagated by the wave equation. This is closely related to the absence of a maximum principle for the wave equation (see Exercise 1.2). The natural function spaces to consider are in fact the H^k spaces, whose norms *are* propagated by the equations[4].

Notice that the support of the function grows at most linearly in time. This is related to the finite speed of propagation. Finally, we have stated a result which is *global in t*, i.e. we do not restrict to a time interval $(-T, T)$. Obviously our result implies similar results on S_T.

Exercise 1.2. Let $\mathbb{R}^3_* := \mathbb{R}^3 \setminus \{0\}$, $S_{*,T} := \mathbb{R}^3_* \times (-T, T)$ and $|x| = r$. You may assume the result that if $u = u(r, t)$ is radial, we have

$$\Delta u(|x|, t) = \Delta u(r, t) = \frac{\partial^2 u}{\partial r^2}(r, t) + \frac{2}{r}\frac{\partial u}{\partial r}(r, t)$$

a) Suppose $u(x, t) = \frac{1}{r} v(r, t)$ for some function v. Show that u solves the wave equation on $\mathbb{R}^3_* \times (0, T)$ if and only if v satisfies the one-dimensional wave equation

$$-\frac{\partial^2 v}{\partial t^2} + \frac{\partial^2 v}{\partial r^2} = 0$$

on $(0, \infty) \times (-T, T)$.

b) Suppose $f, g \in C^2_c(\mathbb{R})$. Deduce that

$$u(x, t) = \frac{f(r + t)}{r} + \frac{g(r - t)}{r}$$

is a solution of the wave equation on $S_{*,T}$ which vanishes for large $|x|$.

c) Show that if $f \in C^3_c(\mathbb{R})$ is an odd function (i.e. $f(s) = -f(-s)$ for all s) then

$$u(x, t) = \frac{f(r + t) + f(r - t)}{2r}$$

extends as a C^2 function which solves the wave equation on S_T, with

$$u(0, t) = f'(t).$$

[4]You will have met these spaces if you took MA4A2: Advanced PDEs. Otherwise you may ignore this comment.

*d) By considering a suitable sequence of functions f, or otherwise, deduce that there exists no constant C independent of u such that the estimate

$$\sup_{S_T} (|u| + |u_t|) \leq C \sup_{\Sigma_0} (|u| + |u_t|)$$

holds for all solutions $u \in C^2(S_T)$ of the wave equation which vanish for large $|x|$.

1.3 Minkowski Spacetime

In the Theory of PDE course, we constructed solutions to the Cauchy problem for the heat equation on $\mathbb{R}^3$:

$$\begin{cases} u_t = \Delta u & \text{in } (0, T) \times \mathbb{R}^3, \\ u = u_0 & \text{on } \{0\} \times \mathbb{R}^3. \end{cases}$$

Physically, the function $u(t, \cdot)$ corresponds to the temperature at time t of an infinite uniform body whose initial temperature is given by the function u_0. In order to write the heat equation out in full, we pick an orthonormal basis for $\mathbb{R}^3$, say $\{\boldsymbol{e}_i\}_{i=1,2,3}$. Such a choice is called a frame. Once we have chosen a frame, we can define the partial derivatives in the $\boldsymbol{e}_i$ directions. For any function $f \in C^1(\mathbb{R}^3)$, we write:

$$\frac{\partial f}{\partial x^i}(\boldsymbol{x}) = \frac{d}{ds} f(\boldsymbol{x} + s\boldsymbol{e}_i)\bigg|_{s=0}.$$

The Laplacian acting on a function u is then given by

$$\Delta f(\boldsymbol{x}) = g^{ij} \frac{\partial^2 f}{\partial x^i \partial x^j}(\boldsymbol{x}) = g^{ij} \frac{\partial}{\partial x^i} \left[\frac{\partial f}{\partial x^j} \right] (\boldsymbol{x}).$$

Here, I am using summation convention, so there is an implicit summation over $i, j = 1, 2, 3$. I have also introduced a new tensor, the Euclidean cometric tensor, whose components are given by:

$$g^{ij} = \begin{cases} 1 & i = j, \\ 0 & i \neq j. \end{cases}$$

As a result, when we perform the implicit summation, the only terms that survive are the diagonal ones and we have:

$$\Delta f(\boldsymbol{x}) = \frac{\partial^2 f}{\partial x^1 \partial x^1}(\boldsymbol{x}) + \frac{\partial^2 f}{\partial x^2 \partial x^2}(\boldsymbol{x}) + \frac{\partial^2 f}{\partial x^3 \partial x^3}(\boldsymbol{x}) \tag{1.2}$$

Now, in the physical explanation of what u represents, I made no reference to any particular frame. The heat equation (1.2), however, makes explicit reference to the coordinates x^i associated to the frame. To reconcile these two facts, we can show that if we change our orthonormal basis, we must find that the heat equation is unchanged.

Suppose $\boldsymbol{e}'_i$ is another basis for $\mathbb{R}^3$. We must have that

$$\boldsymbol{e}_i = \boldsymbol{e}'_j \Lambda^j{}_i$$

for some real numbers $\Lambda^j{}_i$. Now, let us consider how changing basis affects the partial derivatives of a function $f : \mathbb{R}^3 \to \mathbb{R}$. Starting from the definition of the partial derivative, we have

$$\begin{aligned} \frac{\partial f}{\partial x^i}(\boldsymbol{x}) &= \frac{d}{ds} f(\boldsymbol{x} + s\boldsymbol{e}_i)\bigg|_{s=0} \\ &= \frac{d}{ds} f\left(\boldsymbol{x} + s\boldsymbol{e}'_j \Lambda^j{}_i\right)\bigg|_{s=0} \\ &= \Lambda^j{}_i \frac{d}{ds} f\left(\boldsymbol{x} + t\boldsymbol{e}'_j\right)\bigg|_{s=0} \\ &= \Lambda^j{}_i \frac{\partial f}{\partial x'^j}(\boldsymbol{x}) \end{aligned} \tag{1.3}$$

where we have used the chain rule to go from the second to the third line. Since $\Lambda^i{}_j$ don't depend on $\boldsymbol{x}$ we can differentiate again and we find

$$\frac{\partial^2 f}{\partial x^i \partial x^j}(\boldsymbol{x}) = \Lambda^k{}_i \Lambda^l{}_j \frac{\partial^2 f}{\partial x'^k \partial x'^l}(\boldsymbol{x})$$

We see then that the heat equation has the same form with respect to the basis $\{\boldsymbol{e}'_i\}$ as it did for the basis $\{\boldsymbol{e}_i\}$ if and only if

$$\Lambda^k{}_i g^{ij} \Lambda^l{}_j = g^{kl}. \tag{1.4}$$

Thinking of $\Lambda^i{}_j$ and g^{ij} as the components of two matrices, this is equivalent to the matrix equation:

$$\Lambda\Lambda^T = I,$$

which we recognise as the condition that $\Lambda \in O(3)$, i.e. Λ is orthogonal. In other words, the heat equation takes the same form with respect to any orthonormal frame for $\mathbb{R}^3$. This reflects a physical symmetry of the underlying problem: the invariance under rotations and reflections of a uniform body filling all of space.

We could in fact have taken the form of the heat equation to *define* the orthonormal frames. In this approach, we would say that a basis for $\mathbb{R}^3$ is orthonormal if and only if the heat equation takes the form

$$\frac{\partial u}{\partial t}(\boldsymbol{x}) = \frac{\partial^2 u}{\partial x^1 \partial x^1}(\boldsymbol{x}) + \frac{\partial^2 u}{\partial x^2 \partial x^2}(\boldsymbol{x}) + \frac{\partial^2 u}{\partial x^3 \partial x^3}(\boldsymbol{x})$$

with respect to this basis. In principle, we could hope to make physical measurements to determine the form of the heat equation, so this connects a mathematical construction (an orthonormal frame) to the physics of heat propagation.

The wave equation and the Lorentz group

Now we come to the equation we are actually interested in, the wave equation. The aim is, in a similar way to our brief discussion above, to look at the symmetries of the wave equation. The first thing we do is to drop the distinction between time and space.

Trivially, we can consider a point $(t, \boldsymbol{x}) \in \mathbb{R} \times \mathbb{R}^3$ as belonging to $\mathbb{R}^4$, which we wish to think of as the space-time manifold (what this means will become clearer as the course progresses). We introduce a basis for spacetime:

$$\vec{e}_0 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad \vec{e}_1 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \quad \vec{e}_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \quad \vec{e}_3 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

Defining $x^0 := t$, we can write any point in $\mathbb{R}^4$ as

$$\vec{X} := (x^0, \boldsymbol{x}) = x^\mu \vec{e}_\mu$$

where we take the convention that greek indices μ, ν, etc. are summed over $0, \ldots, 3$. In order to write the wave equation concisely, we introduce a $(0,2)-$tensor with components $\eta_{\mu\nu}$ called the metric tensor, given by:

$$\eta_{\mu\nu} = \begin{cases} -1 & \mu = \nu = 0 \\ 1 & \mu = \nu = 1, 2, \text{ or } 3, \\ 0 & \mu \neq \nu. \end{cases} \tag{1.5}$$

Notice that $\{\vec{e}_\mu\}$ is an orthonormal basis for η. The metric tensor has an inverse, a $(2,0)-$tensor with components $\eta^{\mu\nu}$ satisfying

$$\eta^{\mu\sigma}\eta_{\sigma\nu} = \delta^\mu{}_\nu.$$

Here $\delta^\mu{}_\nu$ is the usual Kronecker delta, defined by

$$\delta^\mu{}_\nu = \begin{cases} 1 & \mu = \nu \\ 0 & \mu \neq \nu. \end{cases}$$

We can then write the wave equation very concisely as

$$-u_{tt} + \Delta u = \eta^{\mu\nu} \frac{\partial^2 u}{\partial x^\mu \partial x^\nu} = 0.$$

We'll sometimes use the notation

$$\Box u = \eta^{\mu\nu} \frac{\partial^2 u}{\partial x^\mu \partial x^\nu},$$

in the same way that Δ is used for the Laplacian. The operator $\Box$ is known as the *d'Alembertian* or *wave operator.*

Definition 1. We define the Minkowski spacetime, $\mathbb{E}^{1,3}$, to be $\mathbb{R}^4$ equipped with the metic tensor $\eta \in T^0{}_2(\mathbb{R}^4)$.

1.3.1 Lorentz transformations

As we did for the heat equation above, it is instructive to consider what happens when we choose another basis (or frame) for $\mathbb{R}^4$. Suppose that $\{\vec{e}_\mu{}'\}$ is a new basis for $\mathbb{R}^4$. We of course have

$$\vec{e}_\mu = \vec{e}_\nu{}' \Lambda^\nu{}_\mu,$$

for a set of real numbers $\Lambda^\nu{}_\mu$. The same calculation as (1.3) shows that for any f which is differentiable in a neighbourhood of $X \in \mathbb{E}^{1,3}$ we have

$$\frac{\partial f}{\partial x^\mu}(\vec{X}) = \Lambda^\nu{}_\mu \frac{\partial f}{\partial x'^\nu}(\vec{X}).$$

This means that the numbers

$$\frac{\partial f}{\partial x^\mu}(\vec{X})$$

transform like a $(0,1)-$tensor under a change of basis. It's convenient to write this as

$$\nabla_\mu f := \frac{\partial f}{\partial x^\mu},$$

so that the wave equation can be written

$$\nabla^\mu \nabla_\mu u = \eta^{\mu\nu} \nabla_\nu \nabla_\mu u = \nabla_\nu \eta^{\mu\nu} \nabla_\mu u = \nabla_\mu \nabla^\mu u = 0.$$

It is standard to suppress the argument to keep the notation clean, but it can be replaced if necessary.

Since the $\Lambda^\nu{}_\mu$ are constant, we see that the form of the wave equation is preserved if

$$\eta^{\mu\nu} = \Lambda^\mu{}_\sigma \eta^{\sigma\tau} \Lambda^\nu{}_\tau$$

Thinking of $\Lambda^\mu{}_\nu$ and $\eta_{\mu\nu}$ as matrices, we can write this condition as:

$$\Lambda \eta^{-1} \Lambda^T = \eta^{-1}. \tag{1.6}$$

Definition 2. 1. A basis with respect to which η takes the form (1.5) is called an *inertial frame*. Equivalently, an inertial frame is one for which the wave equation takes the form

$$-\frac{\partial^2 u}{\partial x^0 \partial x^0}(\vec{X}) + \frac{\partial^2 u}{\partial x^1 \partial x^1}(\vec{X}) + \frac{\partial^2 u}{\partial x^2 \partial x^2}(\vec{X}) + \frac{\partial^2 u}{\partial x^3 \partial x^3}(\vec{X}) = 0$$

2. A matrix Λ satisfying (1.6) is said to be a *Lorentz transformation*, and represents a transformation between inertial frames. The set of all Lorentz transformations $O(1,3)$ forms a group under matrix multiplication, called the *Lorentz group*.

The connection between an inertial frame and the wave equation allows us, in principle, to determine whether a frame is inertial or not by making physical measurements of some quantity obeying the wave equation.

Lemma 1.2. *The following conditions are equivalent to (1.6):*

$$\Lambda^{-1} = \eta^{-1}\Lambda^T\eta.$$

and

$$\Lambda^T\eta\Lambda = \eta.$$

Proof. Multiplying (1.6) on the right by η, we have

$$\Lambda\eta^{-1}\Lambda^T = \eta^{-1} \iff \Lambda\left(\eta^{-1}\Lambda^T\eta\right) = I \iff \Lambda^{-1} = \eta^{-1}\Lambda^T\eta.$$

since Λ is a square matrix. For the second condition, multiply the first on the right by Λ and on the left by η to see

$$\Lambda^{-1} = \eta^{-1}\Lambda^T\eta \iff \eta = \eta\eta^{-1}\Lambda^T\eta\Lambda = \Lambda^T\eta\Lambda.$$

□

Since the three conditions are equivalent, we can take any of them to define the Lorentz group.

Example 1. The Lorentz group contains the orthogonal group $O(3)$ as a subgroup. To see this, consider the set of matrices of the form

$$\Lambda = \begin{pmatrix} 1 & 0_{1\times4} \\ 0_{4\times1} & R_{3\times3} \end{pmatrix}$$

where $R_{3\times3} \in O(3)$. It is straightforward to verify that $\Lambda \in O(1,3)$.

The Lorentz group is a *Lie group*[5]. We can think of the Lorentz group as a Lie subgroup of $GL(4)$ (since by Lemma 1.2 Lorentz transformations are invertible).

Theorem 1.5. *The Lorentz group is a Lie subgroup of* $GL(4)$*, and its Lie algebra, denoted by* $\mathfrak{o}(1,3)$ *is the set of matrices* ℓ *satisfying*

$$\ell\eta^{-1} + \eta^{-1}\ell^T = 0.$$

Proof. Consider the map $f : GL(4) \to Sym(4\times4)$ given by

$$f(A) = A\eta^{-1}A^T.$$

Clearly $O(1,3)$ is the preimage of η^{-1}, so it will be enough to show that η^{-1} is a regular point of the map f. To see this, we calculate

$$\begin{aligned} df|_A &= (dA)\eta^{-1}A^T + A\eta^{-1}(dA)^T \\ &= (dA)A^{-1}\eta^{-1} + \eta^{-1}\left[(dA)A^{-1}\right]^T \end{aligned}$$

[5] Don't worry if you've not met these before. All it means is that in addition to having a group structure, $O(1,3)$ can be made into a manifold, in such a way that the group operations are continuous.

for $A \in f^{-1}(\eta^{-1})$. To show that the derivative map is surjective, we must show that for any point $A \in f^{-1}(\eta^{-1})$ and any symmetric matrix $S \in Sym(4 \times 4)$ we can find a matrix B such that

$$BA^{-1}\eta^{-1} + \eta^{-1}\left[BA^{-1}\right]^T = S$$

For this, we can simply take $B = \frac{1}{2}S\eta A$. Thus η^{-1} is a regular point of f, and we deduce that $O(1,3)$ is a closed subgroup of $GL(4)$ and hence is a Lie subgroup. To complete the proof, we recall that the Lie algebra of a matrix Lie group is simply the tangent space at the identity, i.e. $\mathfrak{o}(1,3) = T_I O(1,3)$. From the regular value theorem we know that $T_I O(1,3) = \text{Ker } df|_I$, which corresponds precisely to those matrices ℓ satisfying

$$\ell\eta^{-1} + \eta^{-1}\ell^T = 0.$$

□

The group $O(1,3)$ has four connected components. The connected component containing the identity is a subgroup of $O(1,3)$ called the *proper orthochronous Lorentz group* and is denoted $SO^+(1,3)$. There are two important Lorentz transformations which do not belong to $SO^+(1,3)$. These have matrices[6]:

$$\mathbb{T} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}; \qquad \mathbb{P} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}.$$

Using T, P we can write $O(1,3)$ as the disjoint union of connected cosets:

$$O(1,3) = SO^+(1,3) \cup \mathbb{T}\left(SO^+(1,3)\right) \cup \mathbb{P}\left(SO^+(1,3)\right) \cup \mathbb{PT}\left(SO^+(1,3)\right).$$

We've thus reduced the problem of understanding $O(1,3)$ to the problem of understanding $SO^+(1,3)$. Any element of $SO^+(1,3)$ may be written as

$$\Lambda = e^{\ell} = \sum_{n=0}^{\infty} \frac{\ell^n}{n!}$$

for some $\ell \in \mathfrak{o}(1,3)$, where the exponential here is the matrix exponential.

Example 2. We can take

$$\ell = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

[6] Check that these satisfy (1.6)

We check that

$$\ell\eta^{-1}+\eta^{-1}\ell^T = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} + \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$
$$= \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} = 0.$$

Clearly if $\ell \in \mathfrak{o}(1,3)$, then so is $-s\ell$. The matrix exponential of $-s\ell$ is straightforward to calculate using the fact that

$$\ell^2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

and $\ell^3 = \ell$. We find

$$\begin{aligned} e^{-s\ell} &= \sum_{n=0}^{\infty}\frac{(-s)^n\ell^n}{n!} = I - \ell\sum_{n=0}^{\infty}\frac{s^{2n+1}}{(2n+1)!} + \ell^2\sum_{n=1}^{\infty}\frac{s^{2n}}{(2n)!} \\ &= I - \ell\sinh s + \ell^2(\cosh s - 1) \\ &= \begin{pmatrix} \cosh s & -\sinh s & 0 & 0 \\ -\sinh s & \cosh s & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \end{aligned}$$

This transformation is called a boost in the x direction of rapidity s.

Let us see how the boost of rapidity s changes our coordinates. Recall that our transformation relates the old basis to the new basis by

$$\vec{e}_\mu = \vec{e}_\nu{}'\Lambda^\nu{}_\mu,$$

Suppose we have a point $\vec{X} \in \mathbb{R}^4$ with coordinates x^μ relative to the old basis and x'^μ relative to the new basis. We have:

$$\vec{X} = x^\mu\vec{e}_\mu = x^\mu\Lambda^\nu{}_\mu\vec{e}_\nu{}' = x'^\nu\vec{e}_\nu{}',$$

so that

$$x'^\nu = \Lambda^\nu{}_\mu x^\mu.$$

Thus we have:

$$\begin{aligned} x'^0 &= x^0\cosh s - x^1\sinh s, \\ x'^1 &= -x^0\sinh s + x^1\cosh s, \\ x'^2 &= x^2, \\ x'^3 &= x^2. \end{aligned}$$

We can take this to mean that the frame defined by $\vec{e}_\mu{}'$ is moving at a speed $v = \tanh s$ along the positive x-direction, relative to the frame defined by $\vec{e}_\mu$. Notice that $|v| < 1$. Thus two inertial frames related by a boost of this kind cannot have a relative speed faster than the speed of light.

Classifying vectors

The metric η allows us to define an inner product for vectors in Minkowski space. We define

$$\eta\left(\vec{X}, \vec{Y}\right) = x^\mu y^\mu \eta_{\mu\nu}.$$

By construction, this is invariant under a Lorentz transformation to a new frame.

Exercise 1.3. Show that if $\Lambda^\nu{}_\mu$ is the matrix of a Lorentz transformation and

$$x'^\nu = \Lambda^\nu{}_\mu x^\mu, \qquad y'^\nu = \Lambda^\nu{}_\mu y^\mu.$$

then

$$x'^\mu y'^\mu \eta_{\mu\nu} = x^\mu y^\mu \eta_{\mu\nu}.$$

In particular, for any vector $\vec{X}$ in Minkowski space, we have an invariant quantity $\eta\left(\vec{X}, \vec{X}\right)$ which is the same no matter which inertial frame we calculate it in. Unlike a traditional inner product, this quantity does not need to be positive. We classify non-zero vectors according to the sign.

Definition 3. A non-zero vector $\vec{X} \in \mathbb{E}^{1,3}$ is

i) *Timelike* if $\eta\left(\vec{X}, \vec{X}\right) < 0$,

ii) *Null*, or *lightlike* if $\eta\left(\vec{X}, \vec{X}\right) = 0$,

iii) *Spacelike* if $\eta\left(\vec{X}, \vec{X}\right) > 0$.

A vector is *causal* if it is timelike or null.

We say that a C^1-curve $\gamma : (a, b) \to \mathbb{E}^{1,3}$ is timelike/null/spacelike/causal if its tangent vector $\dot{\gamma}(s)$ is timelike/null/spacelike/causal for *every* $s \in (a, b)$.

You should think of a future directed timelike vector as representing the instantaneous velocity of a particle travelling at less than the speed of light, while a future directed null vector represents the instantaneous velocity of a particle travelling *at* the speed of light. A timelike curve should be thought of as an allowable trajectory for a massive particle.

Since a Lorentz transformation does not change the quantity $\eta\left(\vec{X}, \vec{X}\right)$, we see that if a vector is timelike/null/spacelike with respect to one frame it is timelike/null/spacelike with respect to *every* frame. Similarly, the condition of being past/future directed is independent of frame. In Figure 1.1 we show the surfaces $\eta\left(\vec{X}, \vec{X}\right) = \pm 1, 0$, which consist of vectors of constant Minkowski norm.

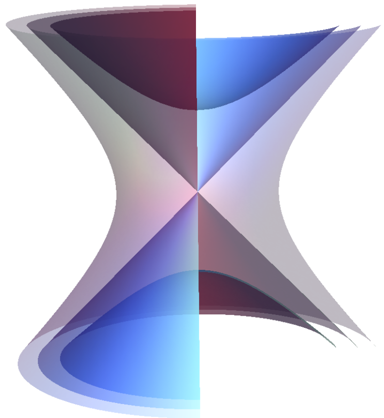

Figure 1.1 The surfaces $\eta\left(\vec{X},\vec{X}\right)=1$ (one sheeted hyperboloid) $\eta\left(\vec{X},\vec{X}\right)=-1$ (two sheeted hyperboloid) and $\eta\left(\vec{X},\vec{X}\right)=0$ (cone)

Exercise 1.4. Using $\mathbb{P},\mathbb{T}$ and the transformations of Examples 1, 2 or otherwise:

a) Suppose that $\vec{X}\in\mathbb{E}^{1,3}$ is a unit timelike vector, i.e. $\eta\left(\vec{X},\vec{X}\right)=-1$. Show that there exists an inertial frame $\{\vec{e}_\mu\}_{\mu=0,\ldots,3}$, such that writing $\vec{X}=x^\mu\vec{e}_\mu$, we have

$$x^\mu=(1,0,0,0).$$

Deduce that if $\vec{X}$ is timelike and $\vec{Y}\neq 0$ satisfies $\eta\left(\vec{X},\vec{Y}\right)=0$, then $\vec{Y}$ is spacelike.

b) Suppose that $\vec{X}\in\mathbb{E}^{1,3}$ is a null vector, i.e. $\eta\left(\vec{X},\vec{X}\right)=0$. Show that there exists an inertial frame $\{\vec{e}_\mu\}_{\mu=0,\ldots,3}$ such that writing $\vec{X}=x^\mu\vec{e}_\mu$, we have

$$x^\mu=\lambda(1,1,0,0).$$

for some $\lambda>0$. Deduce that if $\vec{X}$ is null and $\vec{Y}\neq 0$ satisfies $\eta\left(\vec{X},\vec{Y}\right)=0$, then either $\vec{Y}$ is either spacelike or parallel to $\vec{X}$.

c) Suppose that $X\in\mathbb{E}^{1,3}$ is a unit spacelike vector, i.e. $\eta\left(\vec{X},\vec{X}\right)=1$. Show that there exists an inertial frame $\{\vec{e}_\mu\}_{\mu=0,\ldots,3}$ compatible with the time orientation such that writing $\vec{X}=x^\mu\boldsymbol{e}_\mu$, we have

$$x^\mu=(0,1,0,0).$$

Deduce that if $\vec{X}$ is spacelike and $\vec{Y}\neq 0$ satisfies $\eta\left(\vec{X},\vec{Y}\right)=0$, then $\vec{Y}$ can be timelike, null or spacelike.

It is also useful to classify some surfaces[7] according to their causal properties. We define

[7]for us a surface has three dimensions

Definition 4. We say that a surface $\Sigma \subset \mathbb{E}^{1,3}$ is

i.) *Timelike* if at every point in Σ there is a timelike tangent vector.

ii.) *Null* if at every point in Σ there is a null tangent vector, but no timelike tangent vector.

iii.) *Spacelike* if every vector tangent to Σ is spacelike.

We have the following useful result, which can be proven by choosing a suitable orthonormal basis.

Lemma 1.3. *i.) If Σ is timelike, then locally there exists a spacelike vector field which is normal to the tangent plane (with respect to η).*

ii.) If Σ is spacelike, then locally there exists a timelike vector field which is normal to the tangent plane (with respect to η).

1.3.2 Causal geometry

We can define a relation between timelike vectors $\vec{T}_1$, $\vec{T}_2$ by:

$$\vec{T}_1 \sim \vec{T}_2 \qquad \Longleftrightarrow \qquad \eta(\vec{T}_1, \vec{T}_2) < 0. \tag{1.7}$$

Exercise 1.5. Show that the relation $\sim$ between timelike vectors defined in (1.7) is an equivalence relation. [Hint: the final part of Exercise 1.4 *a*) may be useful]

Definition 5. A *time orientation* for the Minkowski spacetime is an equivalence class of timelike vectors $[\vec{T}]_\sim$. We say that a causal vector, X, is *future directed* if $\eta(\vec{X}, \vec{T}) < 0$ and *past directed* if $\eta(\vec{X}, \vec{T}) > 0$ for any $\vec{T} \in [\vec{T}]_\sim$. We say that an inertial frame $\{\vec{e}_\mu\}_{\mu=0,\ldots,3}$ is compatible with the time orientation $\vec{T}$ if $\vec{e}_0$ is future directed.

Suppose an inertial frame $\{\vec{e}_\mu\}_{\mu=0,\ldots,3}$ is compatible with the time orientation $\vec{T}$, and that $\{\vec{e}_\mu{}'\}_{\mu=0,\ldots,3}$ is related to $\{\vec{e}_\mu\}_{\mu=0,\ldots,3}$ by a Lorentz transformation Λ. We say that Λ preserves the time orientation if $\{\vec{e}_\mu{}'\}_{\mu=0,\ldots,3}$ is compatible with the time orientation. We define the *orthochronous Lorentz group*, $O^+(1,3)$, to be the subgroup of $O(1,3)$ consisting of all of the Lorentz transformations which preserve the time orientation.

Lemma 1.4. *We can decompose the orthochronous Lorentz group as:*

$$O^+(1,3) = SO^+(1,3) \cup \mathbb{P}(SO^+(1,3)).$$

Proof. We first show that transformations in $SO^+(1,3)$ preserve time orientation. Suppose for contradiction that $\{\vec{e}_\mu{}'\}_{\mu=0,\ldots,3}$ is not compatible with the time orientation $[\vec{T}]_\sim$, so that $\eta(\vec{e}_0{}', \vec{T}) \geq 0$ and that the bases are related by

$$\vec{e}_\mu = \vec{e}_\nu{}' \Lambda^\nu{}_\mu,$$

with $\Lambda \in SO^+(1,3)$. This implies that there exists a continuous map $\Gamma : [0,1] \to O(1,3)$ with $\Gamma(0) = I, \Gamma(1) = \Lambda$. Consider the map $f : [0,1] \to \mathbb{R}$ given by

$$f(s) = \eta\left(\vec{e}_\nu{}' \Gamma^\nu{}_0(s), \vec{T}\right).$$

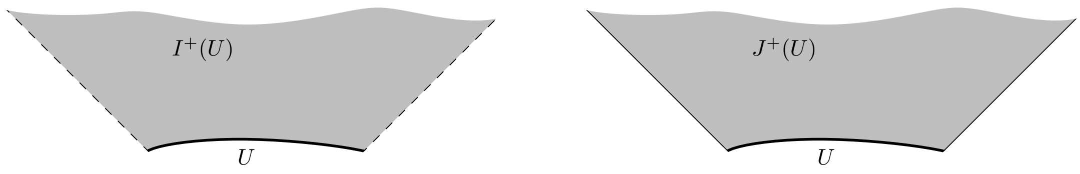

Figure 1.2 The chronological (left) and causal (right) future of a set U

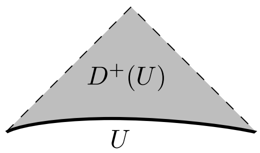

Figure 1.3 The future Cauchy development of a set U

We have that f is continuous, $f(0) < 0$ and $f(1) \geq 0$, so there exists $s_0 \in [0,1]$ such that $f(s_0) = 0$. Since $\Gamma \in SO^+(1,3)$, we know that $\vec{e}_\nu{}'\Gamma^\nu{}_0(s_0)$ is a unit timelike vector, however any non-zero vector orthogonal to the timelike vector $\vec{T}$ must be spacelike, which gives a contradiction.

Finally, we observe that $\mathbb{P}$ preserves the time orientation and $\mathbb{T}$ does not. Together with the decomposition of $O(1,3)$, we obtain the result. □

Obviously, we can define a future directed timelike/causal curve by requiring that the tangent vector is future directed timelike/causal at all points on the curve. This allows us to talk about the future and past of a subset $U \subset \mathbb{E}^{1,3}$. We will define three important sets:

Definition 6. 1. The *chronological future* of $U \subset \mathbb{E}^{1,3}$, denoted $I^+(U)$, is the set of all points $p \in \mathbb{E}^{1,3}$ which can be reached from U by a future directed timelike curve.

2. The *causal future* of $U \subset \mathbb{E}^{1,3}$, denoted $J^+(U)$, is the set of all points $p \in \mathbb{E}^{1,3}$ which can be reached from U by a future directed causal curve.

3. The *future Cauchy development*, or *domain of dependence* of U, denoted $D^+(U)$ is the set of all points $p \in \mathbb{E}^{1,3}$ such that every past inextensible timelike curve through p intersects U.

1.4 Lorentz geometry and the wave equation

1.4.1 Doppler shifts

One simple calculation we can do with the machinery that we've built up in the last section allows us to explain the phenomenon of the Doppler shift. We know this best from the field of acoustics: the engine of a speeding car sounds higher pitched as it approaches us than it does as it drives away. The Doppler shift also occurs for light waves. Light coming from an object moving towards us is 'blue-shifted', meaning the frequency of the electromagnetic radiation increases. Light coming from an object moving away is 'red-shifted'. This fact is behind the dreaded radar guns used to enforce speed limits.

To see how the geometrical picture we've built up can help us understand this phenomenon, let us fix an inertial frame $\{\vec{e}_\mu\}_{\mu=0,\ldots,3}$ and consider the following function[8]

$$\begin{array}{rccl} u_{\vec{k}} & : & \mathbb{E}^{1,3} & \to \mathbb{C} \\ & & X & \mapsto e^{i\eta(\vec{X},\vec{k})} = e^{ix^\mu k_\mu} \end{array}$$

where $\vec{k}$ is a constant vector. We have that

$$\nabla_\mu u_{\vec{k}} = ik_\mu u_{\vec{k}}$$

so that

$$\Box u_{\vec{k}} = -k_\mu k^\mu u_{\vec{k}}.$$

If we choose $\vec{k}$ to be a null vector, so that $k_\mu k^\mu = 0$, then u_k is a solution of the wave equation. These are the plane wave solutions. In this frame, writing $x^\mu = (t, \boldsymbol{x})$, we see

$$u_k = e^{-ik^0 t + i\boldsymbol{k}\cdot\boldsymbol{x}}$$

so that solution in this frame is seen to be a wave with frequency k^0 propagating in the direction $\boldsymbol{n} = \frac{\boldsymbol{k}}{k^0}$.

Now, since our solution is written in a manifestly Lorentz invariant form, we know that $u_{\vec{k}}$ is given by the same expression, regardless of which inertial frame we choose. As a result, if we instead choose an inertial frame $\{\vec{e}_\mu{}'\}_{\mu=0,\ldots,3}$, then we would view the solution as a wave with frequency k'^0 propagating in the direction $\boldsymbol{n}' = \frac{\boldsymbol{k}}{k^0}$. Suppose that the frames are related by the boost transformation of Example 2. Then we have that

$$\begin{aligned} k'^0 &= k^0 \cosh s - k^1 \sinh s, \\ k'^1 &= -k^0 \sinh s + k^1 \cosh s, \\ k'^2 &= k^2, \\ k'^3 &= k^2. \end{aligned}$$

Thus we see that viewed in the two different frames, the frequency and the wavenumber in the x^1 direction differ. We can simplify the situation by assuming that $k^2 = k^3 = 0$ and $k^1 = k^0$, so that our boost is in the direction of propagation. Then we have $k'^2 = k'^3 = 0$ and

$$k'^0 = k'^1 = k^0 e^{-s}.$$

[8] If you prefer to work with real solutions of the wave equation, we are always free to take the real part.

Recalling that the rapidity s corresponds to a relative velocity between the two frames of $v = \tanh s$, we deduce that the frequency in the frame $\{\vec{e}_\mu{}'\}_{\mu=0,\ldots,3}$ is given in terms of the frequency in the frame $\{\vec{e}_\mu\}_{\mu=0,\ldots,3}$ by

$$k'^0 = k^0\sqrt{\frac{1-v}{1+v}}$$

Thus, if $v > 0$, an observer at rest in the primed frame will measure a lower frequency relative to an observer at rest in the unprimed frame. This is because the observer in the primed frame is moving in the positive x^0-direction relative to an observer in the unprimed frame, so sees a red shifted wave.

1.4.2 The energy-momentum tensor

Our motivation for introducing the Lorentz group was to understand symmetries of the wave equation. We found that the wave equation looks the same in any inertial frame. A very deep principle of theoretical physics says that symmetries give rise to conserved quantities. In the case of the wave equation, this fact is demonstrated by the existence of a tensor, the energy-momentum tensor, which has some very useful properties.

Definition 7. Suppose $U \subset \mathbb{E}^{1,3}$ is open. Given $\psi \in C^2(U)$, we define a symmetric $(0,2)-$tensor, the *energy momentum tensor* with components given by

$$T_{\mu\nu}[\psi] := \nabla_\mu\psi\nabla_\nu\psi - \frac{1}{2}\eta_{\mu\nu}\nabla_\sigma\psi\nabla^\sigma\psi.$$

Theorem 1.6. *The energy momentum tensor has the following properties:*

1. *We have a formula for the divergence:*

$$\nabla_\mu T^\mu{}_\nu[\psi] = (\Box\psi)\,\nabla_\nu\psi$$

2. *Fix an inertial frame $\{\vec{e}_\mu\}_{\mu=0,\ldots,3}$. The $00-$component of T is the local energy density*

$$T_{00}[\psi] = \frac{1}{2}\left[(\nabla_0\psi)^2 + |\boldsymbol{\nabla}\psi|^2\right]$$

3. *Fix an inertial frame $\{\vec{e}_\mu\}_{\mu=0,\ldots,3}$. If $\vec{V} = V^\mu\vec{e}_\mu$ is any future directed unit timelike vector, then:*

$$V^0\left[(\nabla_0\psi)^2 + |\boldsymbol{\nabla}\psi|^2\right] \geq V^\mu T_{\mu 0}[\psi] \geq \frac{1}{4V^0}\left[(\nabla_0\psi)^2 + |\boldsymbol{\nabla}\psi|^2\right]$$

Proof. 1. We calculate

$$\begin{aligned}
\nabla_\mu T^\mu{}_\nu[\psi] &= \nabla_\mu\left(\nabla^\mu\psi\nabla_\nu\psi - \frac{1}{2}\delta^\mu{}_\nu\nabla_\sigma\psi\nabla^\sigma\psi\right) \\
&= (\nabla_\mu\nabla^\mu\psi)\,\nabla_\nu\psi + \nabla^\mu\psi\nabla_\mu\nabla_\nu\psi - \frac{1}{2}\nabla_\nu\left(\nabla_\mu\psi\nabla^\mu\psi\right) \\
&= (\nabla_\mu\nabla^\mu\psi)\,\nabla_\nu\psi + \nabla^\mu\psi\nabla_\mu\nabla_\nu\psi - (\nabla_\nu\nabla_\mu\psi)\,\nabla^\mu\psi \\
&= (\Box\psi)\,\nabla_\nu\psi.
\end{aligned}$$

2. Again, a straightforward calculation gives us:

$$T_{00}[\psi] = \nabla_0\psi\nabla_0\psi + \frac{1}{2}\left(-\nabla_0\psi\nabla_0\psi + \nabla_i\psi\nabla^i\psi\right)$$
$$= \frac{1}{2}\left[(\nabla_0\psi)^2 + |\boldsymbol{\nabla}\psi|^2\right]$$

3. Since $\vec{V}$ is timelike, future directed, and unit, we have $V^0 > 0$ and

$$(V^0)^2 - |\boldsymbol{V}|^2 = 1.$$

We calculate, similarly to the previous part:

$$V^\mu T_{\mu 0}[\psi] = \nabla_0\psi(V^0\nabla_0\psi + V^i\nabla_i\psi) + \frac{V^0}{2}\left(-\nabla_0\psi\nabla_0\psi + \nabla_i\psi\nabla^i\psi\right).$$

Now, we use the Cauchy-Schwarz inequality and Young's inequality (or the AM-GM inequality) to deduce

$$\left|\nabla_0\psi V^i\nabla_i\psi\right| \le |\nabla_0\psi|\,|\boldsymbol{V}|\,|\boldsymbol{\nabla}\psi| \le \frac{|\boldsymbol{V}|}{2}\left[(\nabla_0\psi)^2 + |\boldsymbol{\nabla}\psi|^2\right]$$

We therefore have

$$\frac{V^0 + |\boldsymbol{V}|}{2}\left((\nabla_0\psi)^2 + |\boldsymbol{\nabla}\psi|^2\right) \ge V^\mu T_{\mu 0}[\psi] \ge \frac{V^0 - |\boldsymbol{V}|}{2}\left((\nabla_0\psi)^2 + |\boldsymbol{\nabla}\psi|^2\right)$$

Now, since $(V^0)^2 - |\boldsymbol{V}|^2 = 1$, we have $V^0 - |\boldsymbol{V}| = (V^0 + |\boldsymbol{V}|)^{-1}$ and $2V^0 > |\boldsymbol{V}| + V^0$, which completes the proof.

□

Remark: Property 2. is sometimes known as the *weak energy condition* and property 3. as the *dominant energy condition.*

A useful Corollary is the following

Corollary 1.7. *Suppose that $\psi \in C^2(U)$ solves the wave equation in U, and let $\vec{V}$ be a C^1 vector field*[9] *on U. Defining the vector field:*

$$^{\vec{V}}\vec{J}[\psi] = V^\nu T_\nu{}^\mu[\psi]\vec{e}_\mu$$

we have

$$\nabla_\mu\left({}^{\vec{V}}J^\mu[\psi]\right) = {}^V\Pi_{\mu\nu}T^{\mu\nu}[\psi]$$

where

$$^{\vec{V}}\Pi_{\mu\nu} = \nabla_{(\mu}V_{\nu)}$$

is the deformation tensor *of $\vec{V}$. $^{\vec{V}}\vec{J}$ is called the* energy current *associated to the vector field $\vec{V}$.*

[9]This means that with respect to a basis for $\mathbb{E}^{1,3}$ we may write $\vec{V} = V^\mu \boldsymbol{e}_\mu$, where $V^\mu \in C^1(U)$.

Proof. We can calculate, using part 1. of Theorem 1.6:

$$\begin{aligned}
\nabla_\mu \left({}^{\vec{V}}J^\mu \right) &= \nabla_\mu \left(V^\nu T_\nu{}^\mu \right) = (\nabla_\mu V^\nu) T_\nu{}^\mu + V^\nu \nabla_\mu \left(T_\nu{}^\mu \right) \\
&= \nabla_\mu V_\nu T^{\mu\nu} + V^\nu (\Box\psi) \nabla_\nu \psi \\
&= \nabla_{(\mu} V_{\nu)} T^{\mu\nu} = {}^{\vec{V}}\Pi_{\mu\nu} T^{\mu\nu}.
\end{aligned}$$

□

This corollary is particularly useful when the deformation tensor vanishes for $\vec{V}$. These vector fields are especially important:

Definition 8. A *Killing vector* is a vector field such that

$${}^{\vec{V}}\Pi = 0.$$

Killing vectors are closely related to symmetries of the spacetime. To explore a bit more fully the Killing vectors, let us pick an orthonormal basis $\{\vec{e}_\mu\}$ for $\mathbb{E}^{1,3}$ and use it to identify each point $X \in \mathbb{E}^{1,3}$ with its components x^μ with respect to this basis. First, let us note that the constant vector field

$$\vec{P}_0 = \vec{e}_0,$$

is a Killing field: in fact we have that $\nabla_\mu V_\nu = 0$ for all μ, ν. Similarly for the constant vector fields

$$\vec{P}_i = \vec{e}_i.$$

Now, suppose we have a matrix $\ell \in \mathfrak{o}(1,3)$, which can be written in components as $\ell^\mu{}_\nu$. Then we can define a vector field on $\mathbb{E}^{1,3}$ by:

$$\vec{V}_\ell = x^\nu \ell^\mu{}_\nu \vec{e}_\mu.$$

I claim this is a Killing field. To see this, first note that $(V_\ell)^\mu = x^\nu \ell^\mu_\nu$, so that

$$\nabla_\mu (V_\ell)_\nu = \eta_{\nu\sigma} \ell^\sigma{}_\mu$$

hence

$$\begin{aligned}
{}^{\vec{V}_\ell}\Pi_{\mu\nu} &= \frac{1}{2} \left(\eta_{\nu\sigma} \ell^\sigma{}_\mu + \eta_{\mu\sigma} \ell^\sigma{}_\nu \right) \\
&= \left[\eta\ell + (\eta\ell)^T \right]_{\nu\mu} \\
&= \left[\eta \left(\ell\eta^{-1} + \eta^{-1}\ell^T \right) \eta \right]_{\nu\mu} = 0.
\end{aligned}$$

In fact, one can show that any Killing field of the Minkowski spacetime is a linear combination of fields of the form $\vec{P}_\mu$ and $\vec{V}_\ell$. All of these vector fields are useful, but for us, the most useful will be the vector field $\vec{P}_0$, which is everywhere timelike. It owes its existence to the time translation symmetry of the Minkowski spacetime. We will make use of this vector field to prove the final result of this section:

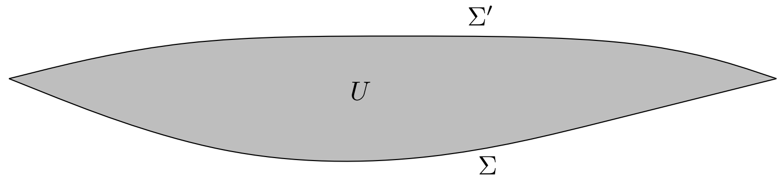

Figure 1.4 The geometry of Theorem 1.8

Theorem 1.8 (Finite speed of propagation). *Suppose that $U \subset \mathbb{E}^{1,3}$ is a open, bounded, set such that the boundary ∂U consists of two smooth compact components Σ, Σ' which are both spacelike, and such that Σ' lies in $J^+(\Sigma)$ (see Figure 1.4). Then if $\psi \in C^2(\overline{U})$ solves the wave equation in U we have the estimate*

$$\int_U \left((\nabla_0 \psi)^2 + |\boldsymbol{\nabla} \psi|^2 \right) dX \leq C \int_\Sigma \left((\nabla_0 \psi)^2 + |\boldsymbol{\nabla} \psi|^2 \right) d\sigma$$

for a constant C which depends only on the domain U. In particular if $\psi, \partial_t \psi$ vanish on Σ, then ψ vanishes throughout U.

Proof. Pick an inertial frame $\{\vec{e}_\mu\}$ for the wave equation, and choose coordinates $\vec{X} = x^\mu \vec{e}_\mu$. Consider the vector field

$$\vec{V} = e^{-x^0} \vec{e}_0$$

We calculate

$$\nabla_\mu V_\nu = \begin{cases} e^{-x^0} & \mu = \nu = 0, \\ 0 & \text{otherwise.} \end{cases}$$

We want to consider the current ${}^{\vec{V}}\vec{J}$. We have

$$\nabla_\mu \left({}^{\vec{V}} J^\mu[\psi] \right) = e^{-x^0} T_{00}$$

We will apply Corollary B.2 (the divergence theorem) to ${}^{\vec{V}}\vec{J}$. We find

$$\int_U \nabla_\mu \left({}^{\vec{V}} \vec{J}^\mu[\psi] \right) dx = \int_\Sigma {}^{\vec{V}} J^\mu[\psi] t_\mu d\sigma - \int_{\Sigma'} {}^{\vec{V}} J^\mu[\psi] t_\mu d\sigma$$

where $\vec{t}$ is the future directed unit normal[10], and $d\sigma$ is the (positive) surface measure. Inserting our expressions for $\nabla_\mu \left({}^{\vec{V}} J^\mu[\psi] \right)$ and ${}^{\vec{V}}\vec{J}$ we have:

$$\int_\Sigma e^{-x^0} T_{0\mu} t^\mu d\sigma = \int_{\Sigma'} e^{-x^0} T_{0\mu} t^\mu d\sigma + \int_U e^{-x^0} T_{00} dx \tag{1.8}$$

Now, since Σ, Σ' are smooth and compact and U is bounded, we have that there exists a constant $c > 1$ such that

$$0 < c^{-1} \leq \sup_{\Sigma \cup \Sigma'} t^0, \ \sup_U |x^0| \leq c < \infty.$$

[10] notice that the signs here are different from what one might expect.

As a result, we have that

$$\int_\Sigma e^{-x^0} T_{0\mu} t^\mu d\sigma \leq c e^c \int_\Sigma \left((\nabla_0 \psi)^2 + |\boldsymbol{\nabla} \psi|^2 \right) d\sigma$$

$$\int_{\Sigma'} e^{-x^0} T_{0\mu} t^\mu d\sigma \geq \frac{e^{-c}}{4c} \int_{\Sigma'} \left((\nabla_0 \psi)^2 + |\boldsymbol{\nabla} \psi|^2 \right) d\sigma \geq 0$$

and

$$\int_U e^{-x^0} T_{00} dX \geq e^{-c} \int_U \left((\nabla_0 \psi)^2 + |\boldsymbol{\nabla} \psi|^2 \right) dx dt$$

Putting these estimates together with (1.8), we have the result. □

Corollary 1.9. *Suppose that $u \in C^\infty(\mathbb{E}^{1,3})$ solves the wave equation, and that*

$$(supp \;\; u|_{x^0=0}) \bigcup (supp \;\; \nabla_0 u|_{x^0=0}) \subset B_R(0)$$

Then

$$(supp \;\; u|_{x^0=T}) \bigcup (supp \;\; \nabla_0 u|_{x^0=T}) \subset B_{R+|T|}(0).$$

This justifies some of the assumptions made in §1.2, where we considered solutions of the wave equation vanishing for large $|\boldsymbol{x}|$ at any fixed time x^0.

Exercise 2.1. Let $U \subset \mathbb{E}^{1,3}$ be open. Define an antisymmetric $(0,2)-$tensor field, F on U with components[11]

$$[F_{\mu\nu}] = \begin{pmatrix} 0 & -E_1 & -E_2 & -E_3 \\ E_1 & 0 & B_3 & -B_2 \\ E_2 & -B_3 & 0 & B_1 \\ E_3 & B_2 & -B_1 & 0 \end{pmatrix}.$$

Where $E_i, B_i \in C^1(U)$. Show that the vacuum Maxwell equations for $\boldsymbol{E}, \boldsymbol{B}$ (with units such that $c^2 = \epsilon_0 \mu_0 = 1$) hold in U if and only if F satisfies the equations

$$\nabla_\mu F^\mu{}_\nu = 0, \qquad \nabla_{[\mu} F_{\nu\sigma]} = 0,$$

where for any $(0,3)-$tensor $A_{\mu\nu\sigma}$ we define:

$$A_{[\mu\nu\sigma]} = \frac{1}{6} \left(A_{\mu\nu\sigma} + A_{\nu\sigma\mu} + A_{\sigma\mu\nu} - A_{\nu\mu\sigma} - A_{\mu\sigma\nu} - A_{\sigma\nu\mu} \right).$$

Exercise 2.2. Suppose that F is as in Exercise 2.1. Fix an inertial frame $\{\vec{e}_\mu\}$. Define

$$T_{\mu\nu}[F] = F_{\mu\sigma} F_\nu{}^\sigma - \frac{1}{4} \eta_{\mu\nu} F_{\sigma\tau} F^{\sigma\tau}.$$

Show that T has the following properties

[11] The convention is that the first index specifies the row, and the second index the column.

a) We have a formula for the divergence:

$$\nabla_\mu T^\mu{}_\nu[F] = \left(\nabla_\mu F^\mu{}_\sigma\right) F_\nu{}^\sigma + \frac{3}{2}\left(\nabla_{[\mu} F_{\nu\sigma]}\right) F^{\mu\sigma}$$

b) The 00−component of T is the local energy density

$$T_{00}[F] = \frac{1}{2}\left[|\boldsymbol{E}|^2 + |\boldsymbol{B}|^2\right]$$

c) If $\vec{V}$ is any future directed unit timelike vector, then:

$$V^0\left[|\boldsymbol{E}|^2 + |\boldsymbol{B}|^2\right] \geq V^\mu T_{\mu 0}[F] \geq \frac{1}{4V^0}\left[|\boldsymbol{E}|^2 + |\boldsymbol{B}|^2\right]$$

Hence, or otherwise, deduce that the electromagnetic field exhibits finite speed of propagation.

Chapter 2

Lorentzian geometry

2.1 The metric and causal geometry

In the previous chapter, we studied Minkowski space, motivated by the wave equation. In that chapter, the metric η (or equivalently, the matrix of coefficients of the wave equation) was constant. We could pick a basis $\{\vec{e}_\mu\}$ of constant vectors (i.e. vectors whose components are independent of x^μ) such that $\eta_{\mu\nu}$ took constant values on $\mathbb{E}^{1,3}$. In this chapter we shall allow the metric to vary from place to place. The correct setting in which to do that is the differentiable manifold.

Manifolds

I'll start by recapping the geometric tools that we shall rely on. We start with an $n-$dimensional C^k-manifold, $\mathcal{M}$, where $k > 3$ can be an integer if the manifold has finite regularity, $k = \infty$ if the manifold is smooth or $k = \omega$ if the manifold is analytic. From the differentiable structure we can define the tangent bundle $T\mathcal{M}$ and the cotangent bundle $T^*\mathcal{M}$, and the bundle of $(p, q)-$tensors, $T^p{}_q\mathcal{M}$. We also define the space $\mathfrak{X}_r(\mathcal{M})$ of C^r-sections of the tangent bundle for $r < k$ (i.e. C^r-smooth vector fields) and the space $\mathfrak{X}^*_r(\mathcal{M})$ of C^r-sections of the co-tangent bundle (i.e. C^r-smooth one-forms). You will need to be familiar with these objects. For a brief introduction to these concepts, see §A.3 or else the MA3H5 Manifolds course.

Definition 9. Let $\mathcal{M}$ be an orientable, $n-$dimensional C^k-manifold and let $r < k$. A C^r-*pseudo-Riemannian metric* on $\mathcal{M}$ is a C^r-smooth $(0, 2)-$tensor field g, such that for each $p \in \mathcal{M}$, $g|_p$ is a symmetric, non-degenerate, billinear form on $T_p\mathcal{M}$. We say that $\mathcal{M}$ equipped with g is a pseudo-Riemannian manifold. The tensor g is called the *metric tensor*.

1.) If $g|_p$ has signature $(+ \ldots +)$ we say that $\mathcal{M}$ equipped with g is a *Riemannian manifold.*

2.) If $g|_p$ has signature $(- + \ldots +)$ we say that $\mathcal{M}$ equipped with g is a *Lorentzian manifold.*

A *spacetime* is a four dimensional Lorentzian manifold.

To unpack this definition a little, suppose that we have a local basis of vector fields $\{e_\mu\}_{i=1,\ldots,n}$ defined on some open set $\mathcal{U}$. In other words, we have $e_\mu \in \mathfrak{X}_{k-1}(\mathcal{U})$, such that $\{e_\mu|_p\}$ span $T_p\mathcal{M}$ for every $p \in \mathcal{U}$. There is a natural basis of one-form fields dual to $\{e_\mu\}_{i=1,\ldots,n}$, which we write as $\{e^\mu\}_{i=1,\ldots,n}$, and is defined by

$$e^\nu[e_\mu] = \delta^\nu{}_\mu$$

With respect to this basis, we write

$$g = g_{\mu\nu} e^\mu \otimes e^\nu.$$

Where for each μ, ν we have $g_{\mu\nu} \in C^r(\mathcal{M}; \mathbb{R})$. The condition that g is symmetric implies that $g_{\mu\nu} = g_{\nu\mu}$, and the non-degeneracy condition implies that the $n \times n$ matrix with components $g_{\mu\nu}$ is invertible. The components of the inverse of this matrix, $g^{\mu\nu}$, which satisfy:

$$g_{\mu\nu} g^{\mu\sigma} = \delta_\nu{}^\sigma.$$

give the components of a symmetric $(2,0)-$tensor, called the co-metic:

$$g^{-1} = g^{\mu\nu} e_\mu \otimes e_\nu.$$

The metric at each point p gives us a bilinear form on $T_p\mathcal{M}$, where we have:

$$g(X, Y) = g_{\mu\nu} e^\mu(X) e^\nu(Y) = g_{\mu\nu} X^\mu Y^\nu$$

For any $X, Y \in T_p\mathcal{M}$ Similarly, the cometric gives a bilinear form on $T_p^*\mathcal{M}$, where:

$$g^{-1}(\omega, \eta) = g^{\mu\nu} \omega(e_\mu) \eta(e_\nu) = g^{\mu\nu} \omega_\mu \eta_\nu.$$

We can use the metric and co-metric to identify $T_p\mathcal{M}$ and $T_p^*\mathcal{M}$. Suppose $X \in T_p\mathcal{M}$, we define the co-vector $X^\flat$ by:

$$X^\flat[Y] = g(X, Y), \qquad \forall \;\; Y \in T_p\mathcal{M}$$

similarly, if $\omega \in T_p^*\mathcal{M}$, we define the vector $\omega^\sharp$ by:

$$\eta[\omega^\sharp] = g^{-1}(\omega, \eta), \qquad \forall \;\; \eta \in T_p^*\mathcal{M}$$

We can check that in a local basis, we have $X^\flat = X^\mu g_{\mu\nu} e^\nu$ and $\omega^\sharp = \omega_\mu g^{\mu\nu} e_\nu$. In this case, we usually write $g_{\mu\nu} X^\mu = X_\nu$ and $\omega_\mu g^{\mu\nu} = \omega^\nu$. In a similar way, we can identify all of the spaces of $(p, q)-$tensors with the same value of $p + q$. Indices are 'raised' with $g^{\mu\nu}$ and 'lowered' with $g_{\mu\nu}$.

Now, to define the signature we use the fact that a non-degenerate bilinear form can be diagonalised. In other words, at each point we can find a basis $\{e_\mu\}$ such that

$$g_{\mu\nu} = g(e_\mu, e_\nu) = \begin{cases} -1 & \mu = \nu < r \\ 1 & \mu = \nu \geq r \\ 0 & \mu \neq \nu \end{cases}$$

for some $r \in \{1, \ldots, n\}$. The number r, which tells us how many positive and how many negative signs appear is independent of the point p at which we diagonalise the metric.

For a Riemannian manifold, the metric is positive definite, each point $p \in \mathcal{M}$ the tangent space $T_p\mathcal{M}$ can be identified with Euclidean space. For a spacetime, at each point $p \in \mathcal{M}$ the tangent space $T_p\mathcal{M}$ can be identified with $\mathbb{E}^{1,3}$. A Lorentzian manifold 'locally looks like Minkowski space', whereas a Riemannian manifold 'locally looks like Euclidean space'.

2.1.1 Examples of pseudo-Riemannian manifolds

We will go through a few examples that will prove useful later in the course.

Example 3. The simplest Riemannian manifold is a $n-$dimensional real vector space V equipped with a positive definite inner product $\langle , \rangle$. A real, finite dimensional vector space is naturally a real analytic manifold, which can be covered by a single coordinate chart as follows. Set $\mathcal{U} = V$. Pick an orthonormal basis $\{\boldsymbol{e}_i\}_{i=1,\ldots n}$ for V with respect to the inner product and define:

$$\begin{array}{rcccl} \varphi & : & \mathcal{U} & \to & \mathbb{R}^n \\ & & \boldsymbol{x} & \mapsto & (x^i)_{i=1,\ldots,n} \end{array}$$

where $\boldsymbol{x} = x^i \boldsymbol{e}_i$.

An element v of $T_{\boldsymbol{x}}V$ can be identified with an element of V in a natural way. Pick a curve $\gamma : (-\epsilon, \epsilon) \to V$ with $\gamma(0) = \boldsymbol{x}$, $\dot{\gamma}(0) = v$. We know from the definition of v that

$$v^i := \left. \frac{d}{dt}\varphi^i(\gamma(t))\right|_{t=0}$$

are independent of the choice of representative curve, where φ^i is the projection of φ onto the i^{th} component, which is a smooth map from V to $\mathbb{R}$. We identify the tangent vector $v \in T_{\boldsymbol{x}}V$ with the vector $\boldsymbol{v} = v^i \boldsymbol{e}_i \in V$. This identification is independent of the choice of basis, so in a natural way we have identified $T_{\boldsymbol{x}}V \simeq V$. Under this identification, we have

$$\frac{\partial}{\partial x^i} \simeq \boldsymbol{e}_i.$$

This identification allows us to define a non-degenerate, symmetric $(0, 2)-$tensor field from the inner product $\langle , \rangle$. Recall that a $(0, 2)-$tensor at a point $\boldsymbol{x}$ is an element of $T^*_{\boldsymbol{x}}V \otimes T^*_{\boldsymbol{x}}V$, or equivalently a bilinear form defined on $T_{\boldsymbol{x}}V$. We define

$$g(v, w) = \langle \boldsymbol{v}, \boldsymbol{w} \rangle$$

where $v, w \in T_{\boldsymbol{x}}V$ are identified with $\boldsymbol{v}, \boldsymbol{w}$.

With respect to the basis $\{\partial_i\}$, we have

$$g\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right) = \langle \boldsymbol{e}_i, \boldsymbol{e}_j \rangle = \delta_{ij}.$$

In order to write g concisely, we can use the dual basis to $\{\partial_i\}$ is denoted $\{dx^i\}$, where

$$dx^i\left(\frac{\partial}{\partial x^j}\right) = \delta^i{}_j.$$

We also introduce the symmetric product of two one-forms, which is an element of $(T_{\boldsymbol{x}})^0{}_2V$, i.e. a $(0,2)-$tensor defined by

$$\omega_1\omega_2 := \frac{1}{2}\left(\omega_1 \otimes \omega_2 + \omega_2 \otimes \omega_1\right)$$

for $\omega_1, \omega_2 \in T^*_{\boldsymbol{x}}V$. This acts on a pair of vectors $v_1, v_2 \in T_{\boldsymbol{x}}V$ as:

$$\omega_1\omega_2(v_1, v_2) = \frac{1}{2}\left[\omega_1(v_1)\omega_2(v_2) + \omega_2(v_1)\omega_1(v_2)\right]$$

With this notation, the metric g can be written:

$$g = \delta_{ij}dx^i dx^j$$

Example 4. A more interesting example of a Riemannian manifold is the unit $n-$sphere:

$$S^n = \left\{\boldsymbol{X} \in \mathbb{R}^{n+1} : \langle \boldsymbol{X}, \boldsymbol{X}\rangle = 1\right\}$$

with $\langle , \rangle$ the canonical inner product on $\mathbb{R}^{n+1}$. The unit $n-$sphere is naturally a real analytic Riemannian manifold (in fact it inherits these properties from $\mathbb{R}^{n+1}$). To cover the sphere, we require two coordinate charts. We'll pick an orthonormal basis $\{\boldsymbol{E}_a\}_{a=1,\ldots n+1}$ for $\mathbb{R}^{n+1}$ and we define

$$\mathcal{U}_\pm = S^n \setminus \{\pm \boldsymbol{E}_{n+1}\}.$$

On each patch, we use stereographic projection to map to $\mathbb{R}^n$, which we endow with the orthonormal basis $\{\boldsymbol{e}_i\}$.

$$\begin{array}{rccl} \varphi_\pm & : & \mathcal{U}_\pm & \to & \mathbb{R}^n \\ & & X^a\boldsymbol{E}_a & \mapsto & \frac{1}{1\mp X^{n+1}}X^i\boldsymbol{e}_i =: \boldsymbol{x}_\pm \end{array}$$

We can invert the transformation by noting that $X^iX_i + (X^{n+1})^2 = 1$ and

$$|\boldsymbol{x}_\pm|^2 = \frac{X^iX^j\delta_{ij}}{(1\mp X^{n+1})^2} = \frac{1-(X^{n+1})^2}{(1\mp X^{n+1})^2} = \frac{1\pm X^{n+1}}{1\mp X^{n+1}}$$

Thus

$$\varphi_\pm^{-1}(\boldsymbol{x}) = \frac{2x^i}{1+|\boldsymbol{x}|^2}\boldsymbol{E}_i \mp \frac{1-|\boldsymbol{x}|^2}{1+|\boldsymbol{x}|^2}\boldsymbol{E}_{n+1} \tag{2.1}$$

From here we conclude:

$$\begin{array}{rccl} \varphi_\mp \circ \varphi_\pm^{-1} & : & \mathbb{R}^n \setminus \{\boldsymbol{0}\} & \to & \mathbb{R}^n \setminus \{\boldsymbol{0}\} \\ & & \boldsymbol{x} & \mapsto & \frac{\boldsymbol{x}}{|\boldsymbol{x}|^2} \end{array}$$

so that the transition functions are indeed real analytic, and S^n is a real analytic manifold.

The tangent space $T_{\boldsymbol{X}}S^n$ can be identified in a natural way with the set:

$$T_{\boldsymbol{X}}S^n \simeq \left\{\boldsymbol{V} \in \mathbb{R}^{n+1} : \langle \boldsymbol{V}, \boldsymbol{X} \rangle = 0\right\}.$$

To see this, consider a tangent vector $V \in T_{\boldsymbol{X}}S^n$ and choose a curve $\gamma : (\epsilon, \epsilon) \to S^n$ with $\gamma(0) = \boldsymbol{X}$, $\dot{\gamma}(0) = V$. Thinking of γ as a curve in $\mathbb{R}^{n+1}$, we can identify its initial tangent vector with $\boldsymbol{V} \in \mathbb{R}^{n+1}$ as in the previous example. We know that the map $f : \mathbb{R}^{n+1} \to \mathbb{R}$ which takes $\boldsymbol{X}$ to $|\boldsymbol{X}|^2$ is smooth, and constant on S^n. By applying the vector $\dot{\gamma}(0)$ to this map we deduce that

$$\begin{aligned} 0 = \dot{\gamma}(0)[f] &= \left.\frac{d}{dt}\left\langle \gamma(t), \gamma(t) \right\rangle\right|_{t=0} \\ &= 2\left\langle \gamma(0), \dot{\gamma}(0) \right\rangle = \left\langle \boldsymbol{V}, \boldsymbol{X} \right\rangle. \end{aligned}$$

This identification of the tangent space of S^n (an abstract manifold construction) with a plane in $\mathbb{R}^{n+1}$ allows us to visualise the two local bases:

$$\left\{\frac{\partial}{\partial x^i_\pm}\right\}_{i=1\ldots n}$$

that are defined via the coordinate charts. Recall that the vector ∂_i at a point $\boldsymbol{X}$ is defined to be the tangent vector to the i^{th} coordinate axis passing through $\boldsymbol{X}$. Accordingly, we can find the bases by differentiating the expression (2.1). We obtain:

$$\begin{aligned} \left.\frac{\partial}{\partial x^i_\pm}\right|_{\boldsymbol{X}=\varphi_\pm^{-1}(\boldsymbol{x})} &\simeq \left(\frac{2}{1+|\boldsymbol{x}|^2}\delta_i{}^j - \frac{4x_i x^j}{\left(1+|\boldsymbol{x}|^2\right)^2}\right)\boldsymbol{E}_j \pm \frac{4x_i}{\left(1+|\boldsymbol{x}|^2\right)^2}\boldsymbol{E}_{n+1} \\ &= \left[(1 \mp X^{n+1})\delta_i{}^j - X_i X^j\right]\boldsymbol{E}_j \pm (1 \mp X^{n+1})X_i \boldsymbol{E}_{n+1} \end{aligned}$$

We see that $\frac{\partial}{\partial x^i_\pm}$ span $T_{\boldsymbol{X}}S^n$ for all $\boldsymbol{X} \in \mathcal{U}_\pm$.

The identification of $T_{\boldsymbol{X}}S^n$ with a subspace of $\mathbb{R}^{n+1}$ gives us a natural way to define a metric on S^n using the inner product of $\mathbb{R}^{n+1}$:

$$g(v, w) = \langle \boldsymbol{V}, \boldsymbol{W} \rangle$$

where $v, w \in T_{\boldsymbol{X}}S^n$ are identified with $\boldsymbol{V}, \boldsymbol{W} \in \mathbb{R}^{n+1}$. We calculate

$$\begin{aligned} g\left(\frac{\partial}{\partial x^i_\pm}, \frac{\partial}{\partial x^j_\pm}\right) &= \left[(1 \mp X^{n+1})\delta_i{}^k - X_i X^k\right]\left[(1 \mp X^{n+1})\delta_{jk} - X_j X_k\right] \\ &\quad + (1 \mp X^{n+1})^2 X_i X^i \\ &= (1 \mp X^{n+1})^2\delta_{ij} - 2X_i X_j(1 \mp X^{n+1}) + X_i X_j X^k X_k \\ &\quad + (1 \mp X^{n+1})^2 X_i X^i \\ &= (1 \mp X^{n+1})^2\delta_{ij} \\ &= \frac{4\delta_{ij}}{\left(1+|\boldsymbol{x}_\pm|^2\right)^2} \end{aligned}$$

We thus have the result that in each stereographic coordinate patch, the standard metric on the unit sphere is given by

$$g = \frac{4\delta_{ij}}{\left(1 + |\boldsymbol{x}_\pm|^2\right)^2} dx^i_\pm dx^j_\pm.$$

Example 5. Let's return to the example of $\mathbb{R}^{n+1}$ equipped with an inner product and an orthonormal basis $\{\boldsymbol{E}_a\}_{a=1,\ldots n+1}$. Suppose we have a point $\boldsymbol{P} \in \mathbb{R}^n \setminus \{\boldsymbol{0}\}$. We can uniquely write $\boldsymbol{P} = r\boldsymbol{X}$, where $r > 0$ is a real number and $\boldsymbol{X} \in S^n$. In this way we can identify $\mathbb{R}^n \setminus \{\boldsymbol{0}\} \simeq (0, \infty) \times S^n$.

We use this observation to cover $\mathbb{R}^n \setminus \{\boldsymbol{0}\}$ with two coordinate patches:

$$\mathcal{U}_\pm = \mathbb{R}^{n+1} \setminus \{\pm r\boldsymbol{E}_{n+1} : r \geq 0\}$$

and define the maps

$$\begin{array}{rcrcl} \varphi_\pm & : & \mathcal{U}_\pm & \to & (0, \infty) \times \mathbb{R}^n \\ & & P^a\boldsymbol{E}_a & \mapsto & \left(|\boldsymbol{P}|, \frac{1}{|\boldsymbol{P}| \mp P^{n+1}} P^i \boldsymbol{e}_i\right). \end{array}$$

In other words, $\varphi_\pm$ maps the point $r\boldsymbol{X}$ to $(r, \boldsymbol{x}_\pm)$, where $\boldsymbol{x}_\pm$ is the stereographic projection of the point $\boldsymbol{X} \in S^n$. The inverse map is given by

$$\varphi_\pm^{-1}(r, \boldsymbol{x}) = r\left(\frac{2x^i}{1 + |\boldsymbol{x}|^2}\boldsymbol{E}_i \mp \frac{1 - |\boldsymbol{x}|^2}{1 + |\boldsymbol{x}|^2}\boldsymbol{E}_{n+1}\right) \tag{2.2}$$

As before, we can identify the tangent vectors with vectors in $\mathbb{R}^{n+1}$. A simple calculation shows that

$$\frac{\partial}{\partial r} \simeq \boldsymbol{X} = \frac{\boldsymbol{P}}{|\boldsymbol{P}|}$$

and

$$\frac{\partial}{\partial n^i_\pm} \simeq r\left\{\left[(1 \mp X^{n+1})\delta_i{}^j - X_i X^j\right]\boldsymbol{E}_j \pm (1 \mp X^{n+1})X_i\boldsymbol{E}_{n+1}\right\}$$

We can calculate

$$g\left(\frac{\partial}{\partial r}, \frac{\partial}{\partial r}\right) = 1, \qquad g\left(\frac{\partial}{\partial r}, \frac{\partial}{\partial n^i_\pm}\right) = 0$$

$$g\left(\frac{\partial}{\partial x^i_\pm}, \frac{\partial}{\partial x^j_\pm}\right) = \frac{4\delta_{ij}}{\left(1 + |\boldsymbol{x}_\pm|^2\right)^2}$$

so that in each patch we have

$$g = dr^2 + r^2\frac{4\delta_{ij}}{\left(1 + |\boldsymbol{x}_\pm|^2\right)^2} dx^i_\pm dx^j_\pm$$

which is the metric on $\mathbb{R}^{n+1} \setminus \{\mathbf{0}\}$ written in spherical polar coordinates, with stereographic coordinates on the sphere. We have shown that $\mathbb{R}^{n+1} \setminus \{\mathbf{0}\}$ is equivalent, as a Riemannian manifold, to $(0, \infty) \times S^n$ endowed with the metric

$$g = dr^2 + r^2 g_{S^n}.$$

Often the fact that this describes $\mathbb{R}^{n+1}$ with the origin removed is elided, and one will talk of the metric on $\mathbb{R}^{n+1}$ written in polar coordinates.

Example 6. Take $\mathcal{M} = \mathbb{R}^4$ with coordinates $(x^0, x^i) = (x^\mu)_{\mu=0,\ldots,3}$. As previously, this gives us a global basis of vector fields $\partial_\mu := \frac{\partial}{\partial x^\mu}$. We can endow $\mathbb{R}^4$ with a Lorentzian metric by:

$$g = \eta_{\mu\nu} dx^\mu dx^\nu = -(dx^0)^2 + (dx^1)^2 + (dx^2)^2 + (dx^3)^2$$

This is of course the Minkowski spacetime. It is clearly Lorentzian, if we take the basis $e_\mu = \partial_\mu$, we have

$$g(e_\mu, e_\nu) = \eta_{\mu\nu}.$$

Example 7. We take $\mathcal{M} = \mathbb{R} \times (0, \infty) \times S^2$. This can be given the structure of a real analytic manifold covered by the coordinate patches

$$\mathcal{U}_\pm = \mathbb{R} \times (0, \infty) \times \mathcal{U}_\pm^{S^2}$$

where $\mathcal{U}_\pm^{S^2}$ are the stereographic coordinate patches on S^2 previously defined. The coordinate charts are given by

$$\begin{array}{rcrcl} \varphi_\pm & : & \mathcal{U}_\pm & \to & \mathbb{R} \times (0, \infty) \times \mathbb{R}^2 \\ & & (t, r, \boldsymbol{X}) & \mapsto & (t, r, \boldsymbol{x}_\pm). \end{array}$$

with the notation as in Example 4. We endow $\mathcal{M}$ with the metric

$$g = -\left(1 - \frac{2m}{r}\right) dt^2 + 2\sqrt{\frac{2m}{r}} dt dr + dr^2 + r^2 \frac{4\delta_{AB} dx_\pm^A dx_\pm^B}{\left(1 + |\boldsymbol{x}_\pm|^2\right)^2},$$

where $A, B = 2, 3$. More concisely, we write

$$g = -\left(1 - \frac{2m}{r}\right) dt^2 + 2\sqrt{\frac{2m}{r}} dt dr + dr^2 + r^2 g_{S^2}.$$

This is a Lorentzian metric, as we can see by exhibiting the following local bases:

$$\begin{aligned} e_0 &= \frac{\partial}{\partial t} - \sqrt{\frac{2m}{r}} \frac{\partial}{\partial r} \\ e_1 &= \frac{\partial}{\partial r} \\ e_A^\pm &= \frac{1 + |\boldsymbol{x}_\pm|^2}{2r} \frac{\partial}{\partial x_\pm^A}. \end{aligned}$$

A short calculation shows that

$$g(e^{\pm}_{\mu}, e^{\pm}_{\nu}) = \eta_{\mu\nu},$$

so that in a neighbourhood of every point in $\mathcal{M}$ we can find a local basis with respect to which the metric is diagonal with entries $(- + ++)$.

This metric is the Painlevé-Gullstrand form of the Schwarzschild spacetime. It represents a spacetime containing a black hole region (we shall see what this means shortly).

2.1.2 Causal geometry for Lorentzian manifolds

We will now update some definitions from the previous Chapter. For the most part, things go through in a very straightforward fashion.

Definition 10. A non-zero vector $X \in T_p\mathcal{M}$ is

i) *Timelike* if $g(X, X) < 0$,

ii) *Null*, or *lightlike* if $g(X, X) = 0$,

iii) *Spacelike* if $g(X, X) > 0$.

The definitions of timelike/null/spacelike curves and surfaces generalise from the Minkowski case in the obvious fashion.

We have to adapt a little the definition of a time orientation in the manifold case.

Definition 11. The manifold $\mathcal{M}$ is *time orientable* if there exists a smooth, nowhere vanishing timelike vector field. For a time orientable manifold, a choice of time orientation is a choice of equivalence class of nowhere vanishing timelike vector fields under the equivalence relation

$$T_1 \sim T_2 \qquad \iff \qquad g(T_1, T_2) < 0 \text{ on } \mathcal{M}.$$

With a time orientation $[T]_\sim$, we can define what it means for a vector $X \in T_p\mathcal{M}$ to be future directed. We say that X is future directed if $g(X, T)|_p < 0$ for any $T \in [T]_\sim$.

For a time oriented manifold we can define the notion chronological and causal future/past and domain of dependence as for the case of Minkowski spacetime.

Example 8. Let us consider the Painlevé-Gullstrand form of the Schwarzschild spacetime, as in Example 7 above. The surface

$$\Sigma_t = \{t\} \times (0, \infty) \times S^2$$

is everywhere spacelike: local bases for the vectors tangent to Σ_t are given by $\{e^{\pm}_1, e^{\pm}_2, e^{\pm}_3\}$, which are all spacelike.

The spacetime is time orientable: the vector field

$$e_0 = \frac{\partial}{\partial t} - \sqrt{\frac{2m}{r}}\frac{\partial}{\partial r}$$

is defined everywhere and timelike, we pick the time orientation defined by $[e_0]_\sim$. In fact, e_0 is the future directed unit normal to Σ_t.

Consider now a future directed causal curve $\gamma : (-\epsilon, \epsilon) \to \mathcal{M}$. We can write

$$\gamma(s) = (t(s), r(s), \boldsymbol{X}(s)) \in \mathbb{R} \times (0, \infty) \times S^2$$

and the tangent vector is given by:

$$\dot\gamma(s) = \dot t(s) \frac{\partial}{\partial t} + \dot r(s) \frac{\partial}{\partial r} + \dot{\boldsymbol{X}}(s)$$

where we understand $\dot{\boldsymbol{X}}(s) \in T_{\boldsymbol{X}(s)} S^2$. We can calculate:

$$g(\dot\gamma(s), e_0) = -\dot t(s) < 0$$

by the fact that γ is future directed. Now consider

$$g(\dot\gamma(s), \dot\gamma(s)) = -\left(1 - \frac{2m}{r(s)}\right) \dot t(s)^2 + 2\sqrt{\frac{2m}{r(s)}} \dot t(s) \dot r(s) + \dot r(s)^2 + r^2 g_{S^2}\left(\dot{\boldsymbol{X}}(s), \dot{\boldsymbol{X}}(s)\right),$$

Now, in order that γ is causal, we must have $g(\dot\gamma(s), \dot\gamma(s)) \leq 0$. Suppose now that $r(s) \leq 2m$ for some s. Then the only way that we can have $g(\dot\gamma(s), \dot\gamma(s)) \leq 0$ is if $\dot r(s) \leq 0$. Thus no future directed causal curve can escape from the region $r \leq 2m$. This region is known as a *black hole region* of the spacetime. We have shown that

$$J^+\left(\{0\} \times (0, 2m] \times S^2\right) \subset [0, \infty) \times (0, 2m] \times S^2$$

In fact, it is fairly straightforward to show that the reverse inclusion holds and that the two sets are equal.

Exercise 2.3. Consider the infinite cylinder $\mathbb{R} \times S^1$ and take as coordinates (x, θ) where $\theta \sim \theta + 2\pi$.

a) Show that when $\mathcal{M}$ is equipped with the Lorentzian metric

$$g = -dx^2 + d\theta^2,$$

it is time-orientable.

b) Now consider the metric

$$g = -\cos\theta dx^2 + 2\sin\theta dx d\theta + \cos\theta d\theta^2$$

i) Show that the vector fields defined for $\theta \in [0, 2\pi)$ by

$$X_0 = \cos\frac{\theta}{2}\partial_x - \sin\frac{\theta}{2}\partial_\theta, \quad X_1 = \sin\frac{\theta}{2}\partial_x + \cos\frac{\theta}{2}\partial_\theta.$$

satisfy

$$g(X_0, X_0) = -1, \quad g(X_0, X_1) = 0, \quad g(X_1, X_1) = 1.$$

Deduce that g is a Lorentzian metric.

ii) Let us denote the point $x = 0, \theta = 0$ by p. Suppose that there exists a nowhere vanishing timelike field T, and without loss of generality assume that $g(X_0, T)|_p < 0$. Show that if $\gamma : [0, 1) \to \mathbb{R} \times [0, 2\pi)$ is any smooth curve with $\gamma(0) = p$ then $g(X_0, T)|_\gamma < 0$.

iii) By considering the curve $\gamma : s \mapsto (0, 2\pi s)$, deduce that $\mathcal{M}$ is not time orientable.

2.2 Affine Connections

We are going to require a means to differentiate vectors (and other tensors). This is a subtle point, and is not completely straightforward. Suppose $\gamma : (-\epsilon, \epsilon) \to \mathcal{M}$ is a curve and $f \in C^1(\mathcal{M})$ is a differentiable function, and we wish to differentiate f along γ. We define

$$\dot{\gamma}(0)[f] = \lim_{h\to 0} \frac{f(\gamma(h)) - f(\gamma(0))}{h}.$$

Now suppose that we have a vector field $X \in \mathfrak{X}_1(\mathcal{M})$. We might try and define the derivative of X along γ by:

$$\dot{\gamma}(0)[X] \stackrel{?}{=} \lim_{h\to 0} \frac{X|_{\gamma(h)} - X|_{\gamma(0)}}{h},$$

however, there is a problem with this. The two vectors whose difference we want to form do not belong to the same vector space: one belongs to $T_{\gamma(h)}\mathcal{M}$ and the other to $T_{\gamma(0)}\mathcal{M}$. We need to decide how to identify the tangent spaces along γ before we can define a quotient such as this. The tool to do this is provided by an 'affine connection'.

Definition 12. Suppose $\mathcal{M}$ is a C^k manifold. A C^r-affine connection ∇, where $r \le k-2$, is a map from $\mathfrak{X}_r(\mathcal{M}) \times \mathfrak{X}_{r+1}(\mathcal{M}) \to \mathfrak{X}_r(\mathcal{M})$ satisfying:

i) For every constant $\lambda \in \mathbb{R}$, $X_i \in \mathfrak{X}_r(\mathcal{M})$, $Y_i \in \mathfrak{X}_{r+1}(\mathcal{M})$ we have

$$\nabla_{X_1}(Y_1 + \lambda Y_2) = \nabla_{X_1} Y_1 + \lambda \nabla_{X_1} Y_2,$$

and

$$\nabla_{X_1 + \lambda X_2} Y_1 = \nabla_{X_1} Y_1 + \lambda \nabla_{X_2} Y_2.$$

ii) If $f \in C^r(\mathcal{M}; \mathbb{R})$, $X \in \mathfrak{X}_r(\mathcal{M})$, $Y \in \mathfrak{X}_{r+1}(\mathcal{M})$, then

$$\nabla_{fX} Y = f \nabla_X Y$$

iii) If $f \in C^{r+1}(\mathcal{M}; \mathbb{R})$, $X \in \mathfrak{X}_r(\mathcal{M})$, $Y \in \mathfrak{X}_{r+1}(\mathcal{M})$, then

$$\nabla_X (fY) = X[f]Y + f\nabla_X Y$$

We call $\nabla_X Y$ the covariant derivative of Y in the direction X. Suppose we are given a basis, $\{e_\mu\}$ for $\mathfrak{X}_{k-1}(\mathcal{U})$, where $\mathcal{U} \subset \mathcal{M}$ is open. Then we have the following result

Lemma 2.1. *i) An C^r-affine connection ∇ is determined on $\mathcal{U}$ by its components with respect to the basis $\{e_\mu\}$, written ${\Gamma^\mu}_{\nu\sigma} \in C^r(\mathcal{U}; \mathbb{R})$ which are defined by*

$$\nabla_{e_\nu} e_\sigma = {\Gamma^\mu}_{\nu\sigma} e_\mu$$

ii) If $\{e'_\mu\}$ is a new basis for $\mathfrak{X}_{k-1}(\mathcal{U})$, related to the old one by:

$$e_\mu = \Lambda^\nu{}_\mu e'_\nu, \qquad \Lambda^\nu{}_\mu \in C^{k-1}(\mathcal{U};\mathbb{R})$$

then the components of ∇ with respect to the old basis are related to those with respect to the new basis by

$$\Gamma^\rho{}_{\nu\mu} = \Lambda^\tau{}_\nu \Lambda^\sigma{}_\mu \Gamma'^\kappa{}_{\tau\sigma}(\Lambda^{-1})^\rho{}_\kappa + \Lambda^\tau{}_\nu e'_\tau \left[\Lambda^\kappa{}_\mu\right](\Lambda^{-1})^\rho{}_\kappa$$

where $(\Lambda^{-1})^\rho{}_\kappa \Lambda^\kappa{}_\tau = \delta^\rho{}_\tau$.

Proof. i) Suppose $X \in \mathfrak{X}_r(\mathcal{M})$, $Y \in \mathfrak{X}_{r+1}(\mathcal{M})$. In $\mathcal{U}$ we can uniquely write $X = X^\mu e_\mu$ and $Y = Y^\mu e_\mu$ for $X^\mu \in C^r(\mathcal{U};\mathbb{R})$, $Y^\mu \in C^{r+1}(\mathcal{U};\mathbb{R})$. Then in $\mathcal{U}$

$$\begin{aligned}\nabla_X Y &= \nabla_{X^\nu e_\nu}\left(Y^\sigma e_\sigma\right)\\ &= X^\nu \nabla_{e_\mu}\left(Y^\sigma e_\sigma\right)\\ &= X^\nu e_\nu(Y^\sigma)e_\sigma + X^\nu Y^\sigma \nabla_{e_\nu} e_\sigma\\ &= \left[X^\nu e_\nu(Y^\mu) + X^\nu Y^\sigma \Gamma^\mu{}_{\nu\sigma}\right] e_\mu\end{aligned}$$

which uniquely determines $\nabla_X Y$.

ii) We calculate

$$\begin{aligned}\nabla_{e_\nu} e_\mu &= \nabla_{\Lambda^\tau{}_\nu e'_\tau}\left(\Lambda^\sigma{}_\mu e'_\sigma\right)\\ &= \Lambda^\tau{}_\nu \Lambda^\sigma{}_\mu \nabla_{e'_\tau} e'_\sigma + \Lambda^\tau{}_\nu e'_\tau\left[\Lambda^\sigma{}_\mu\right] e'_\sigma\\ &= \left(\Lambda^\tau{}_\nu \Lambda^\sigma{}_\mu \Gamma'^\kappa{}_{\tau\sigma} + \Lambda^\tau{}_\nu e'_\tau\left[\Lambda^\kappa{}_\mu\right]\right) e'_\kappa\\ &= \left(\Lambda^\tau{}_\nu \Lambda^\sigma{}_\mu \Gamma'^\kappa{}_{\tau\sigma}(\Lambda^{-1})^\rho{}_\kappa + \Lambda^\tau{}_\nu e'_\tau\left[\Lambda^\kappa{}_\mu\right](\Lambda^{-1})^\rho{}_\kappa\right) e_\rho\end{aligned}$$

where in the last line we use the fact that $e'_\kappa = (\Lambda^{-1})^\rho{}_\kappa e_\rho$.

□

Notice that the second part of the Lemma justifies our restriction $r \leq k-2$, since this is the best regularity for the functions $\Gamma^{\cdot}{}_{\cdot\cdot}$ that is consistent with the regularity of the underlying manifold. It is also important to note that the components $\Gamma^\mu{}_{\nu\sigma}$ do *not* transform as the components of some $(1,2)-$tensor field on the manifold.

A connection allows us to transport a vector along a curve. In order to do this, we require the following result:

Lemma 2.2. *Let $\gamma : [0,1] \to \mathcal{M}$ be a C^2-curve with $\gamma(0) = p$, $\gamma(1) = q$. Suppose that ∇ is a C^r-affine connection, with $r \geq 1$. Given $X_p \in T_p\mathcal{M}$, there is a unique vector field X defined along γ such that*

$$\nabla_{\dot\gamma(t)} X = 0, \qquad X|_{t=0} = X_p. \tag{2.3}$$

We say that X is the parallel transport of X_p along γ.

Proof. Let us suppose that $\gamma([0,1]) \subset \mathcal{U}$, for some open set $\mathcal{U}$ on which we can pick a basis $\{e_\mu\}$ for $\mathfrak{X}_{k-1}(\mathcal{U})$. If this is not the case, we can split γ into a finite number of curves for which this is true. If we write $X = X^\mu e_\mu$ and $X_p = X^\mu_p e_\mu$, then the condition (2.3) becomes

$$\begin{aligned} 0 &= \dot\gamma(t)[X^\mu] + X^\nu \dot\gamma^\sigma \Gamma^\mu{}_{\sigma\nu}, \\ &= \dot X^\mu(t) + X^\nu(t)\dot\gamma^\sigma \Gamma^\mu{}_{\sigma\nu} \end{aligned}$$

together with the initial condition $X^\mu(0) = X^\mu_p(0)$. This is a linear ODE for the coefficients $X^\mu(t)$ with coefficients in C^1, hence a unique solution exists for $t \in [0,1]$. □

Definition 13. A *geodesic* or *auto-parallel curve* with respect to the connection ∇ is a C^2-curve $\gamma : (a,b) \to \mathcal{M}$ satisfying

$$\nabla_{\dot\gamma(t)}\dot\gamma(t) = \lambda(t)\dot\gamma(t)$$

for some $\lambda : (a,b) \to \mathbb{R}$. By a change of parameterisation it is possible to arrange that $\lambda = 0$, in this case, the parameterisation is called an *affine parameterisation.*

In a local coordinate chart, for an affinely parameterised geodesic we can write $\varphi \circ \gamma(t) = (x^\mu(t)) \in \mathbb{R}^n$, so that $\dot\gamma(t) = \dot x^\mu \frac{\partial}{\partial x^\mu}$. Inserting this into the expression for a parallel transported geodesic, we find that the functions $x^\mu(t)$ satisfy:

$$\ddot x^\mu + \dot x^\nu \dot x^\sigma \Gamma^\mu{}_{\nu\sigma} = 0. \tag{2.4}$$

where $\Gamma^\mu{}_{\nu\sigma}$ are the components of the connection with respect to the coordinate induced basis. From basic ODE theory (Picard-Lindelöf theorem), we have:

Lemma 2.3. *Suppose $\mathcal{M}$ is a C^k-manifold, where $k > 2$. Suppose $\mathcal{M}$ is equipped with a C^2-affine connection ∇. Given $p \in \mathcal{M}$, $X_p \in T_p\mathcal{M}$, there exists an $\epsilon > 0$ and an affinely parameterised geodesic curve $\gamma : (-\epsilon,\epsilon) \to \mathcal{M}$ such that*

$$\gamma(0) = p, \qquad \dot\gamma(0) = X_p.$$

Moreover, γ is unique up to extension: If $\gamma' : (-\epsilon',\epsilon') \to \mathcal{M}$ is another such affinely parameterised geodesic where (without loss of generality) $\epsilon' > \epsilon$, then $\gamma'(t) = \gamma(t)$ for $t \in (-\epsilon,\epsilon)$.

Note: we only have local existence in t because the ODE is not linear (the $\Gamma^\mu{}_{\nu\sigma}$ depend on $x(t)$ implicitly).

Exercise 2.4. Consider $\mathcal{M} = \mathbb{R}^3$, with a choice of orthonormal basis $\{\boldsymbol{e}_i\}$ with respect to the Euclidean metric $g_{ij} = \delta_{ij}$. Define a smooth connection on vectors by

$$^{(\tau)}\nabla_{\boldsymbol{e}_i}\boldsymbol{e}_j = \tau\epsilon^k{}_{ij}\boldsymbol{e}_k$$

where ϵ_{ijk} is totally anti-symmetric with $\epsilon_{123} = 1$ and $\tau \in \mathbb{R}$ is a constant. Consider the curve $\gamma : (-1,1) \to \mathbb{R}^3$ given by $\gamma(t) = t\boldsymbol{e}_3$. Show that the vector fields

$$\begin{aligned} \boldsymbol{X}_1 &= \boldsymbol{e}_1\cos(\tau x^3) - \boldsymbol{e}_2\sin(\tau x^3) \\ \boldsymbol{X}_2 &= \boldsymbol{e}_1\sin(\tau x^3) + \boldsymbol{e}_2\cos(\tau x^3) \\ \boldsymbol{X}_3 &= \boldsymbol{e}_3 \end{aligned}$$

are all parallely transported by $^{(\tau)}\nabla$ along γ.

An important property of an affine connection is the *torsion*. This is an anti-symmetric C^r−regular $(1,2)$−tensor field defined by

$$T(X,Y) = \nabla_X Y - \nabla_Y X - [X,Y]$$

where we recall that the commutator of two vector fields is a vector field which acts on functions as:

$$[X,Y]\,(f) = X(Yf) - Y(Xf).$$

If $T(X,Y)$ vanishes, we say that the affine connection ∇ is symmetric.

Exercise 2.5. a) Suppose that $f \in C^1(\mathcal{M};\mathbb{R})$, show that

$$T(fX,Y) = fT(X,Y), \qquad T(X,Y) = -T(Y,X).$$

Deduce that if $\{e_\mu\}$ is locally a basis with $X = X^\mu e_\mu$, $Y = Y^\mu e_\mu$ then:

$$T(X,Y) = T^\sigma{}_{\mu\nu} X^\mu Y^\nu e_\sigma$$

for some C^r−functions $T^\sigma{}_{\mu\nu} := e^\sigma\,[T(e_\mu, e_\nu)]$.

b) Show that for the connection defined in Exercise 2.4, the torsion is given by:

$$T^i{}_{jk} = 2\tau\epsilon^i{}_{jk}.$$

Given a connection ∇, we can define a connection on one-forms. Given $X \in \mathfrak{X}_r(\mathcal{M})$ and $\omega \in \mathfrak{X}^*_{r+1}(\mathcal{M})$, we define $\nabla_X\omega$ by requiring:

$$(\nabla_X\omega)\,[Y] = X\,(\omega[Y]) - \omega\,[\nabla_X Y]\,. \tag{2.5}$$

for any $Y \in \mathfrak{X}_{k-1}(\mathcal{M})$.

Exercise 2.6. Show that if $f \in C^{k-1}(\mathcal{M})$, then (2.5) implies

$$(\nabla_X\omega)\,[fY] = f\,(\nabla_X\omega)\,[Y]$$

for any $Y \in \mathfrak{X}_{k-1}(\mathcal{M})$. Deduce that $\nabla_X\omega \in \mathfrak{X}^*_r(\mathcal{M})$.

We can locally express the covariant derivative of a one-form using the same component functions $\Gamma^\mu{}_{\nu\sigma}$ as we used to express the covariant derivative of a vector:

Lemma 2.4. *Suppose that $\{e_\mu\}$ is a basis for $\mathfrak{X}_{k-1}(\mathcal{U})$, where $\mathcal{U} \subset \mathcal{M}$ is open, and that $\{e^\mu\}$ is a dual basis. Then in $\mathcal{U}$ we have:*

$$\nabla_{e_\nu} e^\mu = -\Gamma^\mu{}_{\nu\sigma} e^\sigma$$

where $\Gamma^\mu{}_{\nu\sigma}$ are defined by

$$\nabla_{e_\nu} e_\sigma = \Gamma^\mu{}_{\nu\sigma} e_\mu.$$

Proof. Consider

$$\begin{aligned}(\nabla_{e_\nu} e^\mu)[e_\tau] &= e_\nu (e^\mu[e_\tau]) - e^\mu [\nabla_{e_\nu} e_\tau] \\ &= e_\nu (\delta^\mu{}_\tau) - e^\mu [\Gamma^\kappa{}_{\nu\tau} e_\kappa] \\ &= 0 - \Gamma^\kappa{}_{\nu\tau} e^\mu [e_\kappa] \\ &= -\Gamma^\mu{}_{\nu\tau}\end{aligned}$$

Since a one-form is completely determined by it's action on a basis of vector fields, this suffices to show the result. □

Finally, we can extend the connection to act on arbitrary tensor fields by requiring that the Leibniz rule applies to tensor products. For example:

$$\nabla_X (Y \otimes \omega) = (\nabla_X Y) \otimes \omega + Y \otimes (\nabla_X \omega),$$

with obvious generalisations to arbitrary tensor products of vector fields and one-forms.

Lemma 2.5. *Suppose S is a C^{r+1} regular (p, q) tensor field, which is written with respect to a local basis $e_\mu \in \mathfrak{X}_{k-1}(\mathcal{U})$ of vector fields as*

$$S = S^{\mu_1,\dots\mu_p}{}_{\nu_1\dots\nu_q} e_{\mu_1} \otimes \dots \otimes e_{\mu_p} \otimes e^{\nu_1} \otimes \dots \otimes e^{\nu_q} \tag{2.6}$$

Where $S^{\mu_1,\dots\mu_p}{}_{\nu_1\dots\nu_q} \in C^{r+1}(\mathcal{U}; \mathbb{R})$. Then $\nabla_X S$ is a C^r regular (p, q) tensor field, with components given in the local basis by:

$$\begin{aligned}(\nabla_X S)^{\mu_1,\dots\mu_p}{}_{\nu_1\dots\nu_q} &= X^\sigma e_\sigma \left[S^{\mu_1\dots\mu_p}{}_{\nu_1,\dots\nu_q}\right] \\ &\quad + X^\sigma \Gamma^{\mu_1}{}_{\sigma\tau} S^{\tau\mu_2,\dots\mu_p}{}_{\nu_1,\dots\nu_q} + \dots + X^\sigma \Gamma^{\mu_j}{}_{\sigma\tau} S^{\mu_1\dots,\mu_{j-1}\tau\mu_{j+1}\dots\mu_p}{}_{\nu_1,\dots\nu_q} + \dots \\ &\quad - X^\sigma \Gamma^\tau{}_{\sigma\nu_1} S^{\mu_1\mu_2\dots\mu_p}{}_{\tau\dots\nu_q} - \dots - X^\sigma \Gamma^\tau{}_{\sigma\nu_j} S^{\mu_1\dots\mu_p}{}_{\nu_1\dots\nu_{j-1}\tau\nu_{j+1}\dots\nu_q} - \dots \\ &=: X^\sigma \nabla_\sigma S^{\mu_1,\dots\mu_p}{}_{\nu_1\dots\nu_q}\end{aligned}$$

We define ∇S to be the $(p, q + 1)-$tensor field with components $\nabla_\sigma S^{\mu_1,\dots\mu_p}{}_{\nu_1\dots\nu_q}$.

Proof. We simply apply the definition of the covariant derivative to (2.6) and then express the right hand side in terms of the basis vectors. □

2.2.1 The Levi-Civita connection

There are many connections on a given manifold. We want to single out a single connection, and it turns out that this is possible when we require the connection to be compatible with our metric and symmetric.

Theorem 2.1 (Fundamental theorem of (pseudo-)Riemannian geometry)**.** *Suppose $\mathcal{M}$ is a $C^{k+1}-$regular pseudo-Riemannian manifold which carries a C^k-regular metric g, where $k \geq 1$. Then there exists a unique symmetric $C^{k-1}-$affine connection, ∇, such that $\nabla g = 0$. This connection is known as the* Levi-Civita connection.

Proof. We prove the statement by an explicit construction. Suppose that $X, Y, Z \in \mathfrak{X}_k(\mathcal{M})$. First note that the symmetry of the connection implies

$$\nabla_X Y - \nabla_Y X = [X, Y] \tag{2.7}$$

Suppose that we have a symmetric connection satisfying $\nabla_X g = 0$, then we have

$$\begin{aligned} X\left[g(Y,Z)\right] &= \nabla_X\left(g(Y,X)\right) \\ &= (\nabla_X g)(X,Y) + g\left(\nabla_X Y, Z\right) + g\left(Y, \nabla_X Z\right) \\ &= g\left(\nabla_X Y, Z\right) + g\left(Y, \nabla_X Z\right) \end{aligned} \tag{2.8}$$

similarly

$$Y\left[g(Z,X)\right] = g\left(\nabla_Y Z, X\right) + g\left(Z, \nabla_Y X\right) \tag{2.9}$$

$$Z\left[g(X,Y)\right] = g\left(\nabla_Z X, Y\right) + g\left(X, \nabla_Z Y\right) \tag{2.10}$$

Taking (2.8)+(2.9)−(2.10), and using (2.7) we have:

$$\begin{aligned} X\left[g(Y,Z)\right] + Y\left[g(Z,X)\right] - Z\left[g(X,Y)\right] = 2g\left(\nabla_X Y, Z\right) &- g\left(\left[X,Y\right], Z\right) \\ &+ g\left(\left[Y,Z\right], X\right) - g\left(\left[Z,X\right], Y\right) \end{aligned}$$

so that, after re-arranging we have

$$\begin{aligned} 2g\left(\nabla_X Y, Z\right) = &X\left[g(Y,Z)\right] + Y\left[g(Z,X)\right] - Z\left[g(X,Y)\right] \\ &+ g\left(\left[X,Y\right], Z\right) + g\left(\left[Z,Y\right], X\right) + g\left(\left[Z,X\right], Y\right) \end{aligned} \tag{2.11}$$

We can calculate the right hand side explicitly in terms of g, X, Y, without needing to know ∇. Since Z is arbitrary, this uniquely determines $\nabla_X Y$ by the non-degeneracy of g. Thus if the connection exists, it is unique.

On the other hand, we can use (2.11) to define a connection. It is straightforward to check that if we do this, that the connection is symmetric and $\nabla g = 0$. □

Exercise 2.7. Consider the connection defined in Exercise 2.4. Show that if we define a Riemannian metric on $\mathbb{R}^3$ by $g(\boldsymbol{v}, \boldsymbol{w}) = \boldsymbol{v} \cdot \boldsymbol{w} = v^i w^j \delta_{ij}$, then the connection ${}^{(\tau)}\nabla$ satisfies

$$\boldsymbol{x}[g(\boldsymbol{y}, \boldsymbol{z})] = g\left({}^{(\tau)}\nabla_{\boldsymbol{x}}\boldsymbol{y}, \boldsymbol{z}\right) + g\left(\boldsymbol{y}, {}^{(\tau)}\nabla_{\boldsymbol{x}}\boldsymbol{z}\right).$$

Deduce that ${}^{(0)}\nabla$ is the Levi-Civita connection of g.

Suppose that we have a local basis of vector fields $\{e_\mu\}$. We know that for each μ, ν, the commutator $[e_\mu, e_\nu]$ is a vector field, so it can be written in terms of the basis vector fields:

$$[e_\mu, e_\nu] = C^\sigma{}_{\mu\nu} e_\sigma,$$

where $C^\sigma{}_{\mu\nu} \in C^{k-1}(\mathcal{M}; \mathbb{R})$ are uniquely determined. Now, let us consider (2.11) applied to the vector fields $X = e_\mu, Y = e_\nu, Z = e_\sigma$, and recall that $g_{\mu\nu} := g(e_\mu, e_\nu)$. We find:

$$\begin{aligned} 2\Gamma^\tau{}_{\mu\nu} g_{\tau\sigma} = &e_\mu(g_{\nu\sigma}) + e_\nu(g_{\sigma\mu}) - e_\sigma(g_{\mu,\nu}) \\ &+ C^\tau{}_{\mu\nu} g_{\tau\sigma} + C^\tau{}_{\sigma\mu} g_{\tau\nu} + C^\tau{}_{\sigma\nu} g_{\tau\mu} \end{aligned}$$

so that multiplying by $g^{\sigma\kappa}$ (the co-metric) we deduce:

$$\Gamma^{\kappa}{}_{\mu\nu} = \frac{1}{2} g^{\sigma\kappa} \Big[e_\mu(g_{\nu\sigma}) + e_\nu(g_{\sigma\mu}) - e_\sigma(g_{\mu\nu}) + C_{\sigma\mu\nu} + C_{\nu\sigma\mu} + C_{\mu\sigma\nu} \Big]$$

This expression simplifies in two useful cases:

1. If $\{e_\mu\}$ is a local coordinate basis, so that $e_\mu = \frac{\partial}{\partial x^\mu}$, then we have $C^\mu{}_{\nu\mu} \equiv 0$ and:
$$\Gamma^{\kappa}{}_{\mu\nu} = \frac{1}{2} g^{\sigma\kappa} \left(\frac{\partial g_{\nu\sigma}}{\partial x^\mu} + \frac{\partial g_{\sigma\mu}}{\partial x^\nu} - \frac{\partial g_{\mu\nu}}{\partial x^\sigma} \right) \tag{2.12}$$
2. If $\{e_\mu\}$ is an orthonormal basis, so that $g_{\mu\nu}$ is a set of constants equal to ± 1 or 0. In this case $e_\mu(g_{\nu\sigma}) \equiv 0$ and we have
$$\Gamma^{\kappa}{}_{\mu\nu} = \frac{1}{2} g^{\sigma\kappa} \left(C_{\sigma\mu\nu} + C_{\nu\sigma\mu} + C_{\mu\sigma\nu} \right)$$

Example 9. Recall the Schwarzschild metric in Painlevé-Gullstrand coordinates (Example 7) is a Lorentzian metric on $\mathcal{M} = \mathbb{R} \times (0, \infty) \times S^2$ given by:

$$g = -\left(1 - \frac{2m}{r}\right) dt^2 + 2\sqrt{\frac{2m}{r}} dt dr + dr^2 + r^2 g_{S^2}.$$

Recall also that we have an orthonormal basis for this matrix given locally by:

$$\begin{aligned} e_0 &= \frac{\partial}{\partial t} - \sqrt{\frac{2m}{r}} \frac{\partial}{\partial r} \\ e_1 &= \frac{\partial}{\partial r} \\ e_A &= \frac{1}{r} b_A, \end{aligned}$$

where $\{b_A\}_{A=2,3}$ is a local orthonormal basis for S^2 (the unit round sphere), such that $[b_A, b_B] = \underline{C}^E{}_{AB} b_E$ for some functions $\underline{C}^E{}_{AB} \in C^\infty(S^2; \mathbb{R})$. We calculate:

$$\begin{aligned} [e_0, e_1] &= -\frac{1}{2r} \sqrt{\frac{2m}{r}} \frac{\partial}{\partial r}, \\ [e_0, e_A] &= \frac{1}{r^2} \sqrt{\frac{2m}{r}} b_A, \\ [e_1, e_A] &= -\frac{1}{r^2} b_A, \\ [e_A, e_B] &= \frac{1}{r^2} \underline{C}^E{}_{AB} b_E. \end{aligned}$$

From here, we deduce that:

$$
\begin{aligned}
C^1{}_{01} &= -C^1{}_{10} = -\frac{1}{2r}\sqrt{\frac{2m}{r}}, \\
C^A{}_{0B} &= -C^A{}_{B0} = \frac{1}{r}\sqrt{\frac{2m}{r}}\delta^A{}_B, \\
C^A{}_{1B} &= -C^A{}_{B1} = -\frac{1}{r}\delta^A{}_B, \\
C^E{}_{AB} &= -C^E{}_{BA} = \frac{1}{r}\underline{C}^E{}_{AB}.
\end{aligned}
$$

and all other components of $C^\mu{}_{\nu\sigma}$ vanish. Since $\{e_\mu\}$ is an orthonormal basis, we can raise and lower indices as in the Minkowski case: we change sign when raising or lowering a 0 index, but not otherwise. We calculate (for example):

$$
\begin{aligned}
\Gamma^1{}_{10} &= \frac{1}{2}\left(C^1{}_{10} - C^0{}_{11} + C^1{}_{10}\right) \\
&= \frac{1}{2r}\sqrt{\frac{2m}{r}}.
\end{aligned}
$$

Similar calculations lead us to conclude that the non-vanishing components of Γ with respect to the basis $\{e_\mu\}$ are:

$$
\begin{aligned}
\Gamma^1{}_{10} &= \frac{1}{2r}\sqrt{\frac{2m}{r}}, & \Gamma^0{}_{11} &= \frac{1}{2r}\sqrt{\frac{2m}{r}}, & \Gamma^A{}_{B0} &= -\frac{1}{r}\sqrt{\frac{2m}{r}}\delta^A{}_B, \\
\Gamma^A{}_{B1} &= \frac{1}{r}\delta^A{}_B, & \Gamma^0{}_{AB} &= -\frac{1}{r}\sqrt{\frac{2m}{r}}\delta_{AB}, & \Gamma^1{}_{AB} &= \frac{1}{r}\delta_{AB}, \\
& & \Gamma^C{}_{AB} &= \frac{1}{r}\underline{\Gamma}^C{}_{AB}, & &
\end{aligned}
$$

where $\underline{\Gamma}^C{}_{AB}$ are the connection coefficients of the Levi-Civita connection of the unit metric on S^2 with respect to the basis $\{b_A\}$.

We can use these coefficients to directly give the action of the Levi-Civita connection of g on the basis $\{e_\mu\}$:

$$
\begin{aligned}
\nabla_{e_0}e_0 &= 0, & \nabla_{e_1}e_0 &= \frac{1}{2r}\sqrt{\frac{2m}{r}}e_1, & \nabla_{e_A}e_0 &= -\frac{1}{r}\sqrt{\frac{2m}{r}}e_A, \\
\nabla_{e_0}e_1 &= 0, & \nabla_{e_1}e_1 &= \frac{1}{2r}\sqrt{\frac{2m}{r}}e_0, & \nabla_{e_A}e_1 &= \frac{1}{r}e_A,
\end{aligned} \tag{2.13}
$$

and

$$
\nabla_{e_0}e_A = 0, \qquad \nabla_{e_1}e_A = 0, \qquad \nabla_{e_A}e_B = -\delta_{AB}\left(\frac{1}{r}\sqrt{\frac{2m}{r}}e_0 + \frac{1}{r}e_1\right) + \frac{1}{r}\underline{\Gamma}^C{}_{AB}e_C.
$$

Exercise(∗). Repeat for yourself the calculations of the previous example.

Definition 14. A *geodesic* of the metric g, is a geodesic of the Levi-Civita connection of g, i.e. a curve $\gamma : (-\epsilon, \epsilon) \to \mathcal{M}$ satisfying

$$\nabla_{\dot{\gamma}(t)}\dot{\gamma}(t) = \lambda(t)\dot{\gamma}(t)$$

For a suitable choice of parameterisation, the function $\lambda(t)$ may be chosen to vanish, in which case we talk of an affinely parameterised geodesic.

The geodesic equations can also be derived from a variational principle. In Riemannian geometry this is especially useful as geodesics can be shown to locally minimise the distance between two points. For Lorentzian geometries, the geodesics typically extremise the length, but do not necessarily minimise it.

Exercise(*) (If you have studied Lagrangian mechanics)**.** Suppose we work in a coordinate chart $\mathcal{U} \subset \mathcal{M}$ in which the metric takes the form:

$$g = g_{\mu\nu}dx^\mu dx^\nu.$$

Show that the affinely parameterised geodesic equations

$$\ddot{x}^\mu + \dot{x}^\nu\dot{x}^\sigma\Gamma^\mu{}_{\nu\sigma} = 0,$$

where

$$\Gamma^\kappa{}_{\mu\nu} = \frac{1}{2}g^{\sigma\kappa}\left(\frac{\partial g_{\nu\sigma}}{\partial x^\mu} + \frac{\partial g_{\sigma\mu}}{\partial x^\nu} - \frac{\partial g_{\mu\nu}}{\partial x^\sigma}\right)$$

can be derived from the Lagrangian:

$$L = \frac{1}{2}g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu.$$

Exercise 3.1. Suppose that S is a $(1,1)-$tensor and R is a $(0,2)-$tensor, both at least C^1-regular, which with respect to a local basis $\{e_\mu\}$ may be written:

$$S = S^\mu{}_\nu e_\mu \otimes e^\nu, \qquad R = R_{\sigma\tau}e^\sigma \otimes e^\tau.$$

Suppose ∇ is the Levi-Civita connection.

i) With the notation of Lemma 2.5, show that:

$$\nabla_\kappa\left(S^\mu{}_\nu R_{\sigma\tau}\right) = \left(\nabla_\kappa S^\mu{}_\nu\right)R_{\sigma\tau} + S^\mu{}_\nu\left(\nabla_\kappa R_{\sigma\tau}\right).$$

ii) how also that

$$\nabla_\kappa\delta^\mu{}_\sigma = 0, \qquad \nabla_\kappa g_{\mu\nu} = 0, \qquad \nabla_\kappa g^{\mu\nu}.$$

Conclude that ∇, thought of as a map from $(p,q)-$tensors to $(p,q+1)-$tensors commutes with contractions and raising and lowering of indices.

2.2.2 The wave equation

Suppose that $\mathcal{M}$ is a C^k-manifold with a $C^{k-1}-$regular Riemannian or Lorentzian metric g, which has an associated $C^{k-2}-$regular Levi-Civita connection ∇. Given $f \in C^k(\mathcal{M};\mathbb{R})$, and $X, Y \in \mathfrak{X}_{k-1}(\mathcal{M})$ we define:

$$H_f(X,Y) := X(Yf) - (\nabla_X Y)\, f.$$

Lemma 2.6. *H_f is a symmetric, $C^{k-2}-$regular $(0,2)-$tensor.*

Proof. To show that H_f is symmetric, we calculate:

$$\begin{aligned} H_f(X,Y) - H_f(Y,X) &= X(Yf) - Y(Xf) - [(\nabla_X Y) - (\nabla_Y X)]\, f \\ &= [X,Y]f - [(\nabla_X Y) - (\nabla_Y X)]\, f \\ &= T(X,Y)f = 0. \end{aligned}$$

To show that H_f is a tensor, we need to show that for any $g \in C^{k-1}(\mathcal{M})$ we have:

$$H_f(gX,Y) = gH_f(X,Y) = H_f(X,gY).$$

By the symmetry we can just show the first equality, which follows from

$$H_f(gX,Y) = (gX)[Yf] - (g\nabla_X Y)f = gH_f(X,Y).$$

The regularity of H_f follows from the assumptions on the connection and on f. □

The tensor H_f is called the Hessian of f and is sometimes written $\nabla\nabla f$, so that if $\{e_\mu\}$ is a local basis, we write:

$$\nabla_\mu \nabla_\nu f = H_f(e_\mu, e_\nu).$$

Note carefully that in general

$$\nabla_\mu \nabla_\nu f \neq \nabla_{e_\mu} \nabla_{e_\nu} f = e_\mu(e_\nu f).$$

Exercise 3.2. In a local coordinate basis, $\{e_\mu = \frac{\partial}{\partial x^\mu}\}$, show that we can write

$$H_f(X,Y) = X^\mu Y^\nu \left(\frac{\partial^2 f}{\partial x^\mu \partial x^\nu} - \Gamma^\sigma{}_{\mu\nu} \frac{\partial f}{\partial x^\sigma} \right),$$

Deduce that for a local coordinate basis

$$\nabla_\mu \nabla_\nu f = \frac{\partial^2 f}{\partial x^\mu \partial x^\nu} - \Gamma^\sigma{}_{\mu\nu} \frac{\partial f}{\partial x^\sigma}$$

Definition 15. Suppose $f \in C^k(\mathcal{M};\mathbb{R})$.

i) If g is *Lorentzian*, we define the wave operator $\Box_g$ acting on f, written $\Box_g f \in C^{k-2}(\mathcal{M};\mathbb{R})$ is defined to be the trace of H_f with respect to the metric g. With respect to a local coordinate basis $\{\partial/\partial x^\mu\}$, this is given by:

$$\Box_g f := g^{\mu\nu} \nabla_\mu \nabla_\nu f = \nabla^\mu \nabla_\mu f.$$

ii) If g is *Riemannian*, we define the Laplace operator Δ_g acting on f, written $\Delta_g f \in C^{k-2}(\mathcal{M};\mathbb{R})$ to be the trace of H_f with respect to the metric g. With respect to a local coordinate basis $\{\partial/\partial x^i\}$, this is given by:

$$\Delta_g f := g^{ij}\nabla_i\nabla_i f = \nabla^i\nabla_i f.$$

In a local coordinate basis, we have

$$\Box_g f = g^{\mu\nu}\frac{\partial^2 f}{\partial x^\mu \partial x^\nu} - g^{\mu\nu}\Gamma^{\sigma}{}_{\mu\nu}\frac{\partial f}{\partial x^\sigma}, \tag{2.14}$$

with a similar expression for the Laplace operator:

$$\Delta_g f = g^{ij}\frac{\partial^2 f}{\partial x^i \partial x^j} - g^{ij}\Gamma^{k}{}_{ij}\frac{\partial f}{\partial x^k}, \tag{2.15}$$

When g is *Lorentzian*, we recognise $\Box_g$ as a hyperbolic differential operator. In the case where g is Riemannian, the operator Δ_g is rather an elliptic differential operator.

It's often useful to have a more concrete expression for this operator. We can find one by recalling that the Levi-Civita connection has components that are given in a local coordinate basis by (2.12):

$$\Gamma^{\kappa}{}_{\mu\nu} = \frac{1}{2}g^{\sigma\kappa}\left(\frac{\partial g_{\nu\sigma}}{\partial x^\mu} + \frac{\partial g_{\sigma\mu}}{\partial x^\nu} - \frac{\partial g_{\mu\nu}}{\partial x^\sigma}\right)$$

Contracting with the co-metric, we have:

$$g^{\mu\nu}\Gamma^{\kappa}{}_{\mu\nu} = g^{\sigma\kappa}\frac{\partial g_{\nu\sigma}}{\partial x^\mu}g^{\mu\nu} - \frac{1}{2}g^{\sigma\kappa}\frac{\partial g_{\mu\nu}}{\partial x^\sigma}g^{\mu\nu}$$

Now, recall the following facts regarding differentiating matrices:

$$\frac{dA^{-1}}{dt}(t) = -A^{-1}\frac{dA}{dt}(t)A^{-1}(t).$$
$$\frac{d}{dt}\det A(t) = [\det A(t)]\,\mathrm{Tr}\left(\frac{dA}{dt}(t)A^{-1}(t)\right).$$

We recognise that we can write

$$g^{\mu\nu}\Gamma^{\kappa}{}_{\mu\nu} = -\frac{\partial}{\partial x^\mu}g^{\mu\kappa} - \frac{g^{\sigma\kappa}}{\sqrt{|g|}}\frac{\partial}{\partial x^\sigma}\sqrt{|g|}$$

where we have defined $|g| := |\det(g_{\mu\nu})|$. Returning to (2.14), we deduce that in local coordinates:

$$\Box_g f = \frac{1}{\sqrt{|g|}}\frac{\partial}{\partial x^\mu}\left(\sqrt{|g|}g^{\mu\nu}\frac{\partial f}{\partial x^\nu}\right).$$

A similar expression of course holds in the case that g is Riemannian:

$$\Delta_g f = \frac{1}{\sqrt{|g|}}\frac{\partial}{\partial x^i}\left(\sqrt{|g|}g^{ij}\frac{\partial f}{\partial x^j}\right).$$

An alternative way to define the Hessian and wave/Laplace operator is instead to recall that given a function $f \in C^r(\mathcal{M};\mathbb{R})$, there is a natural one-form $df \in \mathfrak{X}^*_{r-1}(\mathcal{M})$ defined by:

$$df(X) = Xf,$$

for any $X \in \mathfrak{X}(\mathcal{M})$. With respect to a local basis:

$$df = (e_\mu f)e^\mu = (\nabla_\mu f)e^\mu.$$

Since we have a metric g we can associate df with a vector field, sometimes called $\mathrm{grad}_g f$, which acts on $h \in C^1(\mathcal{M};\mathbb{R})$ by:

$$[\mathrm{grad}_g f](h) = g^{-1}(df, dh).$$

In a local basis:

$$\mathrm{grad}_g f = g^{\mu\nu}(df)_\nu e_\mu = g^{\mu\nu}(e_\nu f)e_\mu$$

To define the wave/Laplace operator, we introduce the divergence of a vector field as follows. If $V \in \mathfrak{X}_r(\mathcal{M})$, then the divergence of V, written $\mathrm{div}_g V \in C^{r-1}(\mathcal{M};\mathbb{R})$ is given by:

$$\mathrm{div}_g V = e^\mu \left[\nabla_{e_\mu} V\right] = \nabla_\mu V^\mu.$$

Clearly, we have

$$\Box_g f = \mathrm{div}_g \left(\mathrm{grad}_g f\right).$$

Exercise 3.3. For $V \in \mathfrak{X}_r(\mathcal{M})$, the divergence of V, written $\mathrm{div}_g V \in C^{r-1}(\mathcal{M};\mathbb{R})$ is defined with respect to a local basis by:

$$\mathrm{div}_g V := \nabla_\mu V^\mu = e^\mu \left(\nabla_{e_\mu} V\right).$$

Show that with respect to a local coordinate basis we have:

$$\mathrm{div}_g V = \frac{1}{\sqrt{|g}} \frac{\partial}{\partial x^\mu} \left(\sqrt{|g|} V^\mu\right).$$

Deduce the expression for the wave/Laplace operator with respect to local coordinates, using:

$$\Box_g f = \mathrm{div}_g \left(\mathrm{grad}_g f\right).$$

We can in fact generalise the definition of the wave/Laplace operator to arbitrary tensors. For example, if S is a C^2-regular $(1,1)-$tensor, we can define (in a local basis):

$$\Box_g S = \left(\nabla_\mu \nabla^\mu S^\nu{}_\tau\right) e_\nu \otimes e^\tau.$$

Example 10. Consider the sphere S^2, with the stereographic coordinate charts $(\mathcal{U}^\pm, \varphi_\pm)$ that we previously defined. Recall that the metric on each patch is given by:

$$g_{S^2} = \Omega^2 \delta_{ij} dx^i dx^j.$$

Where

$$\Omega = \frac{2}{1 + |\boldsymbol{x}|^2}$$

As a matrix, the components g_{ij} are

$$[g_{ij}] = \Omega^2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

so that $|g| = \Omega^4$ and

$$[g^{ij}] = \Omega^{-2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

so that $\sqrt{|g|} g^{ij} = \delta^{ij}$. As a consequence, we have

$$\Delta_{S^2} f = \Omega^{-2} \Delta_{\mathbb{R}^2} f$$

where

$$\Delta_{\mathbb{R}^2} f = \delta^{ij} \frac{\partial^2 f}{\partial x^i x^j}$$

is the standard Laplacian on $\mathbb{R}^2$.

Consider the following integral:

$$I := \int_{|\boldsymbol{x}|<R} f \Delta_{S^2} f \sqrt{|g|} dx = \int_{|\boldsymbol{x}|<R} f \Delta_{\mathbb{R}^2} f \ dx.$$

Now, we can use the standard divergence theorem on $\mathbb{R}^2$ to deduce

$$I = \int_{|\boldsymbol{x}|=R} f \frac{\partial f}{\partial \nu} d\sigma - \int_{|\boldsymbol{x}|<R} \delta^{ij} \frac{\partial f}{\partial x^i} \frac{\partial f}{\partial x^j} dx.$$

I claim that if f is a smooth function on S^2, then the first term vanishes as $R \to \infty$, and we can conclude:

$$\int_{\mathbb{R}^2} f \Delta_{S^2} f \, \Omega^2 dx = - \int_{\mathbb{R}^2} \delta^{ij} \frac{\partial f}{\partial x^i} \frac{\partial f}{\partial x^j} dx.$$

We immediately conclude that a harmonic function on S^2 (i.e. a function $f \in C^2(S^2; \mathbb{R})$ satisfying $\Delta_{S^2} f = 0$) must necessarily be constant.

To check the vanishing of the boundary term, by considering the transition function between the two stereographic patches, it's simple to check that if f is a smooth function on S^2 which is given in one stereographic patch by $f(\boldsymbol{x})$, then the function $\tilde{f}(\boldsymbol{x}) = f\left(\frac{\boldsymbol{x}}{|\boldsymbol{x}|^2}\right)$ should be smooth at the origin. Consider

$$\frac{\partial \tilde{f}}{\partial x^i}(\boldsymbol{x}) = \frac{\delta^{ij} |\boldsymbol{x}|^2 - 2x^i x^j}{|\boldsymbol{x}|^4} \frac{\partial f}{\partial x^j} \left(\frac{\boldsymbol{x}}{|\boldsymbol{x}|^2} \right)$$

Setting $\boldsymbol{y} = \frac{\boldsymbol{x}}{|\boldsymbol{x}|^2}$, and noting that $\frac{\partial \tilde{f}}{\partial x^i}(\boldsymbol{x})$ is bounded as $\boldsymbol{x} \to 0$, we conclude that for large $|y|$, we have

$$\left| \frac{\partial f}{\partial x^i}(\boldsymbol{y}) \right| \leq \frac{C}{|\boldsymbol{y}^2|}$$

from which the result easily follows.

Exercise 3.4. Consider the sphere $S^2 = \{\boldsymbol{x} \in \mathbb{R}^3 : |\boldsymbol{x}| = 1\}$, and set

$$\mathcal{U} = S^2 \setminus \{(-\sqrt{1-t^2}, 0, t) : t \in [-1, 1]\},$$

i.e. the sphere with a line of longitude from north to south pole removed. We define a coordinate chart $(\mathcal{U}, \varphi)$, where φ is defined in terms of its inverse by:

$$\begin{array}{rccc} \varphi^{-1} & : & (0,\pi)\times(-\pi,\pi) & \to & \mathcal{U} \\ & & (\theta,\phi) & \mapsto & \begin{pmatrix} \sin\theta\cos\phi \\ \sin\theta\sin\phi \\ \cos\theta \end{pmatrix}. \end{array}$$

a) Show that with the standard identification of tangent vectors in S^2 with vectors in $\mathbb{R}^3$ that we have

$$\left.\frac{\partial}{\partial\theta}\right|_{\varphi^{-1}(\theta,\phi)} \simeq \boldsymbol{e}_\theta := \begin{pmatrix} \cos\theta\cos\phi \\ \cos\theta\sin\phi \\ \sin\theta \end{pmatrix}, \qquad \left.\frac{\partial}{\partial\phi}\right|_{\varphi^{-1}(\theta,\phi)} \simeq \boldsymbol{e}_\phi := \begin{pmatrix} -\sin\theta\sin\phi \\ \sin\theta\cos\phi \\ 0 \end{pmatrix}.$$

b) Show that

$$\begin{aligned} \langle \boldsymbol{e}_\theta, \boldsymbol{e}_\theta\rangle|_{\varphi^{-1}(\theta,\phi)} &= 1, \\ \langle \boldsymbol{e}_\theta, \boldsymbol{e}_\phi\rangle|_{\varphi^{-1}(\theta,\phi)} &= 0, \\ \langle \boldsymbol{e}_\phi, \boldsymbol{e}_\phi\rangle|_{\varphi^{-1}(\theta,\phi)} &= \sin^2\theta. \end{aligned}$$

where $\langle,\rangle$ is the standard inner product on $\mathbb{R}^3$.

c) Deduce that in the θ, ϕ coordinates:

$$g_{S^2} = d\theta^2 + \sin^2\theta d\phi^2.$$

d) Show that in these coordinates:

$$\Delta_{g_{S^2}} f = \frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial f}{\partial\theta}\right) + \frac{1}{\sin^2\theta}\frac{\partial^2 f}{\partial\phi^2}$$

e) Suppose f is a smooth function on S^2. Show that:

$$\int_{\theta=0}^{\pi}\int_{\phi=-\pi}^{\pi} f\Delta_{g_{S^2}} f \sin\theta d\theta d\phi = -\int_{\theta=0}^{\pi}\int_{\phi=-\pi}^{\pi}\left[\left(\frac{\partial f}{\partial\theta}\right)^2 + \frac{1}{\sin^2\theta}\left(\frac{\partial f}{\partial\phi}\right)^2\right]\sin\theta d\theta d\phi$$

Example 11. A more interesting example is furnished by the Painlevé-Gullstrand form of the Schwarzschild Black Hole. Recall that in a coordinate patch we have

$$g = -\left(1 - \frac{2m}{r}\right)dt^2 + 2\sqrt{\frac{2m}{r}}dtdr + dr^2 + r^2\Omega^2\delta_{ij}dx^i dx^j,$$

with Ω as above. Writing the components of the metric as a matrix, we have

$$[g_{\mu\nu}] = \begin{pmatrix} -\left(1 - \frac{2m}{r}\right) & \sqrt{\frac{2m}{r}} & 0 & 0 \\ \sqrt{\frac{2m}{r}} & 1 & 0 & 0 \\ 0 & 0 & \Omega^2 & 0 \\ 0 & 0 & 0 & \Omega^2 \end{pmatrix}$$

so that $|g| = r^4\Omega^4$, and

$$[g^{\mu\nu}] = \begin{pmatrix} -1 & \sqrt{\frac{2m}{r}} & 0 & 0 \\ \sqrt{\frac{2m}{r}} & \left(1 - \frac{2m}{r}\right) & 0 & 0 \\ 0 & 0 & \Omega^{-2} & 0 \\ 0 & 0 & 0 & \Omega^{-2} \end{pmatrix}$$

We find

$$\Box_g f = -\frac{\partial^2 f}{\partial t^2} + \sqrt{\frac{2m}{r}}\frac{\partial^2 f}{\partial t \partial r} + \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\sqrt{\frac{2m}{r}}\frac{\partial f}{\partial t}\right) + \frac{1}{r^2}\frac{\partial}{\partial r}\left([r^2 - 2mr]\frac{\partial f}{\partial r}\right) + \frac{1}{r^2}\Delta_{S^2} f. \tag{2.16}$$

In the same way that studying the wave equation in Minkowski space helped us to understand the causal properties (such as signal propagation) for that spacetime, solutions of the $\Box_g u = 0$ are of great interest in studying the spacetime $(\mathcal{M}, g)$. We will give a taste of results in this direction, but the study of wave equations on Lorentzian manifolds is a large field of study with many fascinating recent advances.

Theorem 2.2. *Let $(\mathcal{M}, g)$ be the Painlevé-Gullstrand form of the Schwarzschild spacetime, as in Example 7 above. Suppose that $f \in C^2(\mathcal{M}; \mathbb{R})$ vanishes for sufficiently large r at every fixed time t. Define:*

$$E_f[t] := \frac{1}{2}\int_{(2m,\infty)\times\mathbb{R}^2} \left[\left(\frac{\partial f}{\partial t}\right)^2 + \left(1 - \frac{2m}{r}\right)\left(\frac{\partial f}{\partial r}\right)^2 + |\not\nabla f|^2\right] r^2\Omega^2 dr dx$$

where

$$\Omega^2 := \frac{4}{\left(1 + |\boldsymbol{x}|^2\right)^2}, \qquad \not\nabla_A f = \frac{1}{r\Omega}\frac{\partial}{\partial x^A}.$$

Then we have

$$\frac{dE_f}{dt} + 4m^2 \int_{\mathbb{R}^2} \left.\frac{\partial f}{\partial t}\right|_{r=2m}^2 \Omega^2 dx = -\int_{(2m,\infty)\times\mathbb{R}^2} \Box_g f \frac{\partial f}{\partial t}\, r^2\Omega^2 dr dx.$$

Proof. We multiply the expression (2.16) by:

$$-\frac{\partial f}{\partial t} r^2 \Omega^2$$

and integrate over $(2m, \infty) \times \mathbb{R}^2$ with the measure $drdx$. The first term gives

$$\int_{(2m,\infty)\times\mathbb{R}^2} \frac{\partial^2 f}{\partial t^2}\frac{\partial f}{\partial t} r^2 \Omega^2 drdx = \frac{d}{dt}\frac{1}{2}\int_{(2m,\infty)\times\mathbb{R}^2} \left(\frac{\partial f}{\partial t}\right)^2 r^2\Omega^2 drdx$$

The second term we integrate by parts in r to give:

$$\begin{aligned} -\int_{(2m,\infty)\times\mathbb{R}^2} \sqrt{\frac{2m}{r}}\frac{\partial^2 f}{\partial t\partial r}\frac{\partial f}{\partial t} r^2\Omega^2 drdx = &+\int_{(2m,\infty)\times\mathbb{R}^2} \frac{\partial}{\partial r}\left(r^2\sqrt{\frac{2m}{r}}\frac{\partial f}{\partial t}\right)\frac{\partial f}{\partial t}\Omega^2 drdx \\ &+ 4m^2\int_{\mathbb{R}^2} \left.\frac{\partial f}{\partial t}\right|^2_{r=2m} \Omega^2 dx \end{aligned}$$

where we use the fact that f vanishes for sufficiently large r. This term will cancel with the other mixed derivative term, leaving only the boundary term. Next consider the term involving two radial derivatives. We have

$$\begin{aligned} -\int_{(2m,\infty)\times\mathbb{R}^2} &\frac{1}{r^2}\frac{\partial}{\partial r}\left([r^2-2mr]\frac{\partial f}{\partial r}\right)\frac{\partial f}{\partial t} r^2\Omega^2 drdx \\ &= \int_{(2m,\infty)\times\mathbb{R}^2} [r^2-2mr]\frac{\partial^2 f}{\partial t\partial r}\frac{\partial f}{\partial r}\Omega^2 drdx \\ &= \frac{d}{dt}\frac{1}{2}\int_{(2m,\infty)\times\mathbb{R}^2}\left[1-\frac{2m}{r}\right]\left(\frac{\partial f}{\partial r}\right)^2 r^2\Omega^2 drdx \end{aligned}$$

Where we use the fact that f vanishes for sufficiently large r, and that the factor $r^2 - 2mr$ vanishes at $r = 2m$ to discard boundary terms. Finally, we use the result

$$\int_{\mathbb{R}^2} f\Delta_{S^2} f\, \Omega^2 dx = -\int_{\mathbb{R}^2} \delta^{ij}\frac{\partial f}{\partial x^i}\frac{\partial f}{\partial x^j} dx.$$

to integrate the last term by parts in x and we're done. □

Corollary 2.3. *Suppose $u_1, u_2 \in C^2(\mathcal{M})$ are two solutions of the wave equation on the Schwarzschild spacetime, vanishing for sufficiently large r at each fixed t, such that:*

$$u_1 = u_2, \qquad \frac{\partial}{\partial t}u_1 = \frac{\partial}{\partial t}u_2, \quad \text{on } \{0\}\times(2m,\infty)\times S^2$$

Then $u_1 = u_2$ in the region $\mathcal{R} = [0,\infty)\times(2m,\infty)\times S^2$.

Proof. We apply Theorem 2.2 to $u_1 - u_2$. We immediately conclude that $E_{u_1-u_2}[t]$ is monotone decreasing and positive, however, it is initially zero thus $E_{u_1-u_2}[t] = 0$ for $t \geq 0$. We conclude that $u_1 - u_2 \equiv 0$ in the region $\mathcal{R}$. □

In fact, this result follows from a more abstract result that says that a solution of the wave equation on a Lorentzian manifold is determined in $D^+(\Sigma)$ by its Cauchy data on Σ.

Proposition 1. *Let $(\mathcal{M}, g)$ be a smooth, orientable, time orientable, Lorentzian manifold. Suppose that $\mathcal{U} \subset \mathcal{M}$ is a open set with compact closure, whose boundary consists of two smooth compact components: Σ, Σ', which are both spacelike and such that $\mathcal{U} \subset D^+(\Sigma)$. Suppose that $\psi \in C^2(\overline{\mathcal{U}})$ solves the equation:*

$$\Box_g \psi = 0 \tag{2.17}$$

in $\mathcal{U}$. Then if

$$\psi|_\Sigma = 0, \qquad N_\Sigma \psi|_\Sigma = 0,$$

Where N_Σ is the future unit normal of Σ, we have that $\psi = 0$ in $\mathcal{U}$.

The proof of this result will mirror (and extend) the proof of Theorem 1.8, the analogous result for Minkowski space, and will require us to reintroduce some machinery from the Minkowski proof, now adapted to the manifold case.

Definition 16. Suppose $\mathcal{U} \subset \mathcal{M}$ is open. Given $\psi \in C^2(\mathcal{U})$, we define a symmetric $(0, 2)-$tensor, the *energy momentum tensor* by

$$T[\psi] := d\psi \otimes d\psi - \frac{1}{2} |d\psi|_g \, g$$

or, with respect to a local basis $\{e_\mu\}$:

$$T[\psi] = \left((e_\mu \psi)(e_\nu \psi) - \frac{1}{2} g_{\mu\nu} (e_\sigma \psi)(e_\tau \psi) g^{\sigma\tau} \right) e^\mu \otimes e^\nu \tag{2.18}$$

$$= \left(\nabla_\mu \psi \nabla_\nu \psi - \frac{1}{2} g_{\mu\nu} \nabla_\sigma \psi \nabla^\sigma \psi \right) e^\mu \otimes e^\nu \tag{2.19}$$

Exercise 3.5. In the case that $(\mathcal{M}, g)$ is the Minkowski spacetime and that $\{e_\mu\}$ is an inertial frame, show that (2.18) agrees with the previous defninition of the energy momentum tensor.

Theorem 2.4. *The energy momentum tensor has the following properties:*

1. *We have a formula for the divergence:*

$$div_g T[\psi] := \nabla_\mu T^\mu{}_\nu[\psi] e^\nu = (\Box_g \psi) \, d\psi$$

2. *Suppose $V \in T_p\mathcal{M}$ is a unit timelike vector. Then at p, we have $T[\psi](V, V) \geq 0$, with equality iff $d\psi$ vanishes at p. If $\{e_\mu\}$ is any basis for $T_p\mathcal{M}$, there exists a constant $C > 0$ depending on V and $\{e_\mu\}$ such that:*

$$\frac{1}{C} \sum_\mu (e_\mu \psi)^2 \leq T[\psi](V, V) \leq C \sum_\mu (e_\mu \psi)^2 .$$

3. *If $W \in T_p\mathcal{M}$ is an unit timelike vector, then:*

$$\frac{1}{4 |g(V, W)|} T[\psi](V, V) \leq T[\psi](V, W) \leq |g(V, W)| \, T[\psi](V, V).$$

Proof. 1. We work in a local basis. We have:

$$\begin{aligned}\nabla_\mu T^\mu{}_\nu &= \nabla_\mu \left(\nabla^\mu \psi \nabla_\nu \psi - \frac{1}{2} \delta^\mu{}_\nu \nabla_\sigma \psi \nabla^\sigma \psi \right) \\ &= \Box_g \psi \nabla_\nu \psi + \nabla^\mu \psi \nabla_\mu \nabla_\nu \psi - \nabla_\nu \nabla_\sigma \psi \nabla^\sigma \psi \\ &= \Box_g \psi \nabla_\nu \psi\end{aligned}$$

2, 3. The proofs of the weak and dominant energy conditions follow directly from the same proofs in the Minkowski case: We pick an orthonormal basis for $T_p\mathcal{M}$ in which $V = e^0$, and then follow the same calculations as before. Since we work only in the tangent space at a point, all of the same calculations are valid.

□

We shall require a version of the divergence theorem for a Lorentzian manifold:

Lemma 2.7. *Suppose $\mathcal{M}$ is a smooth orientable, time orientable, $(n+1)-$dimensional Lorentzian manifold. Suppose that $V \in \mathfrak{X}_1(\mathcal{M})$ is a vector field and that $\mathcal{U} \subset \mathcal{M}$ is an open set with compact closure, whose boundary $\Sigma = \partial\mathcal{U}$ consists piecewise of smooth embedded submanifolds and can be written as $\Sigma = \Sigma_s \cup \Sigma_t$ where Σ_s is spacelike and Σ_t is timelike. Then we have*

$$\int_{\mathcal{U}} \mathrm{div}_g V dX = \int_{\Sigma_s} g(V, N) d\sigma - \int_{\Sigma_t} g(V, N) d\sigma,$$

where vector N is the unit outwards *normal (with respect to g). The volume measure dX and surface measure $d\sigma$ are positive, and on each coordinate chart are equivalent to the $(n+1)-$dimensional (resp. $n-$dimensional) Lebesgue measure.*

The final result that we shall require is a statement about the causal structure of $D^+(\Sigma)$ for a spacelike Σ:

Lemma 2.8. *Suppose $\mathcal{M}$ is a smooth orientable, time orientable, $(n+1)-$dimensional Lorentzian manifold. Suppose $\Sigma \subset \mathcal{M}$ is a smooth embedded $n-$dimensional spacelike submanifold, with $D^+(\Sigma)$ non-empty. Then there exists a function $t \in C^\infty(D^+(\Sigma); \mathbb{R})$, such that:*

i) $\Sigma = \{t = 0\}$.

ii) The level sets of t are spacelike, equivalently $\mathrm{grad}_g t$ *is everywhere causal.*

iii) t increases along any future directed causal curve, equivalently $-\mathrm{grad}_g t$ *is future directed.*

Such a t is called a *time function* for $D^+(\Sigma)$. The proof that such a function exists is quite subtle and is beyond the scope of the course. It is fairly straightforward to produce a C^0-function which is increasing on any future directed causal curve, and whose level sets are achronal (i.e. no two points on a single level set can be connected by a timelike curve). Showing that this C^0-result can be improved to higher regularity is not trivial.[1]

We now have the pieces we require to prove the proposition.

Proof of Proposition 1. 1. Consider the vector field given in a local basis $\{e_\mu\}$ by:

$$^V J[\psi] = T_\mu{}^\nu[\psi] V^\mu e_\nu$$

If ψ solves the wave equation, we have

$$\mathrm{div}_g \left(^V J[\psi] \right) = T^{\mu\nu}[\psi]\, ^V\Pi_{\mu\nu}$$

where the deformation tensor is given in a local basis by:

$$^V\Pi := \nabla_{(\mu} V_{\nu)} = \frac{1}{2} \left(\nabla_\mu V_\nu + \nabla_\nu V_\mu \right) e^\nu \otimes e^\mu.$$

2. By the assumptions of the proposition together with Lemma 2.8, we have on $D^+(\Sigma)$ a smooth time function t, and a vector X which is everywhere timelike. We define

$$V = e^{-\lambda t} X.$$

We apply the divergence theorem to $^V J[\psi]$ on the region $\mathcal{U}$ to find:

$$\int_{\mathcal{U}} \mathrm{div}_g \left(^V J[\psi] \right) dX = \int_\Sigma g \left(^V (J[\psi], N \right) d\sigma - \int_{\Sigma'} g \left(^V (J[\psi], N \right) d\sigma$$

where N is the future directed unit normal vector.

3. We calculate:

$$^V\Pi_{\mu\nu} = -\lambda e^{-\lambda t} X_{(\mu} \nabla_{\nu)} t + {}^X\Pi_{\mu\nu} e^{-\lambda t}$$

so that locally we have:

$$\mathrm{div}_g \left(^V J[\psi] \right) = -\lambda T_{\mu\nu}[\psi] X^\mu \nabla^\nu t + T^{\mu\nu}[\psi]^X \Pi_{\mu\nu}$$

4. By the fact that $-\nabla^\nu t$ is future directed and timelike, we have that $-T_{\mu\nu}[\psi] X^\mu \nabla^\nu t$ is positive, and controls all derivatives of ψ at each point by part 3 of Theorem 2.4. For sufficiently large λ, making use of the fact that $\mathcal{U}$ has compact closure, we conclude that

$$\mathrm{div}_g \left(^V J[\psi] \right) \geq e^{-\lambda t} T[\psi](X, X) \geq C^{-1} T[\psi](X, X)$$

everywhere in $\mathcal{U}$ for some large finite C, by the compactness of $\overline{\mathcal{U}}$.

---

[1]Those interested can find the C^0-result in the paper:
"Domain of dependence," R. Geroch, J.Math.Phys. **11** (1970) 437-439.
The smooth result is in the paper:
"On Smooth Cauchy hypersurfaces and Geroch's splitting theorem,"paper
A. Bernal, M. Sanchez Commun.Math.Phys. **243** (2003) 461-470

5. Turning to the surface terms, we have
$$g\left({}^V J[\psi], N\right) = T[\psi](V, N)$$
so again using the compactness of Σ, Σ', we have:
$$\int_\Sigma g\left({}^V J[\psi], N\right) d\sigma \leq C \int_\Sigma T[\psi](X, X) d\sigma$$
and
$$\int_{\Sigma'} g\left({}^V J[\psi], N\right) d\sigma \geq C^{-1} \int_{\Sigma'} T[\psi](X, X) d\sigma$$
6. We conclude that there exists a constant C, independent of ψ such that
$$\int_{\mathcal{U}} T[\psi](X, X) dX \leq C \int_\Sigma T[\psi](X, X) d\sigma.$$
Now recall that $T[\psi](X, X)$ vanishes at a point if and only if $d\psi$. We see that if ψ and $N_\Sigma \psi$ vanish on Σ (and hence $d\psi = 0$ on Σ), we must have $df = 0$in $\mathcal{U}$ and hence $\psi = 0$.

□

Notice that in fact we have a stronger statement than in the proposition, we in fact have a quantitative estimate for a solution ψ of the wave equation in terms of initial data. If you attended the advanced PDE course, you will recognise that we have in fact established H^1-control of ψ in the region $\mathcal{U}$ in terms of the H^1-initial data.

2.3 Curvature

In order to consider the dynamical gravitational field, we want to write down some equation satisfied by the metric. There are three main goals we have in arriving at this equation:

1. The equation should be hyperbolic, that is to say it locally has similar properties to the wave equation.
2. The equation should be geometric, it shouldn't depend on a particular choice of basis or coordinate system.
3. The gravitational field should couple to matter in a natural way.

The first condition will ensure that the features we observe in special relativity (such as finite speed of signal propagation) are not affected by the gravitational field. The second condition is a mathematical consequence of the equivalence principle. The third condition, while somewhat vague, is necessary since we expect the gravitational field to interact with matter in the spacetime.

One possible first guess at how we could achieve these goals would be to assume that the metric tensor obeys the wave equation:
$$\Box_g g = F$$

where the wave operator is defined on a symmetric 2−tensor in the obvious fashion $(\Box_g h)_{\mu\nu} = \nabla^\sigma \nabla_\sigma h_{\mu\nu}$ in a local basis, and F should represent some source term for the gravitational field. Unfortunately, it is trivial that $\Box_g g \equiv 0$, so this won't work. It turns out that the correct object to play the role of '$\Box g$' is a term involving the *curvature* of the metric. Before we can write down the Einstein equations then, we need to take a bit of time to define curvatures.

2.3.1 Riemann curvature

Suppose that $\mathcal{M}$ is a C^k−manifold, $k \geq 3$, equipped with a C^r-regular pseudo-Riemannian metric g, where $2 \leq r < k$. The associated C^{r-1}−regular Levi-Civita connection is ∇. Given $X, Y \in \mathfrak{X}_{k-1}(\mathcal{M})$ and a C^s−regular (p,q)−tensor S, with $s \leq r$, we define the C^{s-2}−regular (p,q) tensor $R(X,Y)S$ by:

$$R(X,Y)S := \nabla_X \nabla_Y S - \nabla_Y \nabla_X S - \nabla_{[X,Y]} S. \tag{2.20}$$

Lemma 2.9. *The tensor $R(X,Y)S$ has the following properties:*

i) It is antisymmetric in X, Y:

$$R(X,Y)S = -R(Y,X)S$$

for any $X, Y \in \mathfrak{X}_{k-1}(\mathcal{M})$ and C^s−regular (p,q) tensor S.

ii) It is f−linear in the first (and hence second) slot:

$$R(fX_1 + X_2, Y)S = fR(X_1, Y)S + R(X_2, Y)S,$$

for any $f \in C^1(\mathcal{M}; \mathbb{R}), X_i, Y \in \mathfrak{X}_{k-1}(\mathcal{M})$ and C^s−regular (p,q) tensor S.

iii) It is f−linear in S:

$$R(X,Y)[fS_1 + S_2] = fR(X,Y)S_1 + R(X,Y)S_2,$$

for any $f \in C^2(\mathcal{M}; \mathbb{R}), X, Y \in \mathfrak{X}_{k-1}(\mathcal{M})$ and C^s−regular (p,q) tensors S_i.

iv) The following metric compatibility condition holds:

$$g\left(R(X,Y)Z, W\right) + g\left(Z, R(X,Y)W\right) = 0, \qquad \forall \quad X, Y, Z, W \in \mathfrak{X}_{k-1}(\mathcal{M}; \mathbb{R}).$$

Proof. i) This is immediate by inspection of the definition (2.20).

ii) First note that it is straightforward to see from the definition that

$$R(X_1 + X_2, Y)S = R(X_1, Y)S + R(X_2, Y)S,$$

which follows from the $\mathbb{R}$-linearity of the connection. It remains to show that $R(fX, Y)S = fR(X,Y)S$. For this we calculate

$$\begin{aligned} R(fX,Y)S &= \nabla_{fX} \nabla_Y S - \nabla_Y \nabla_{fX} S - \nabla_{[fX,Y]} S, \\ &= f\nabla_X \nabla_Y S - \nabla_Y \left(f \nabla_X S\right) - \nabla_{f[X,Y] - Y(f)X} S, \\ &= f\nabla_X \nabla_Y S - f\nabla_Y \nabla_X S - Y(f)\nabla_X S - \left(f\nabla_{[X,Y]} S - Y(f)\nabla_X S\right), \\ &= f\left(\nabla_X \nabla_Y S - \nabla_Y \nabla_X S - \nabla_{[X,Y]} S\right), \\ &= fR(X,Y)S. \end{aligned}$$

iii) Again, the $\mathbb{R}-$linearity of the connection quickly gives us that

$$R(X,Y)[S_1+S_2] = R(X,Y)S_1 + R(X,Y)S_2,$$

so it remains to show that $R(X,Y)[fS] = fR(X,Y)S$. We calculate:

$$\begin{aligned}
R(X,Y)[fS] &= \nabla_X\nabla_Y(fS) - \nabla_Y\nabla_X(fS), \\
&= \nabla_X(f\nabla_Y S + Y(f)S) - \nabla_Y(f\nabla_X S + X(f)S) - \nabla_{[X,Y]}(fS), \\
&= f\nabla_X\nabla_Y S + X(f)\nabla_Y S + Y(f)\nabla_X S + X(Yf)S \\
&\quad - (f\nabla_Y\nabla_X S + Y(f)\nabla_X S + X(f)\nabla_Y S + Y(Xf)S) \\
&\quad - f\nabla_{[X,Y]}S - ([X,Y]f)\, S \\
&= fR(X,Y)S + (X(Yf) - Y(Xf) - [X,Y]f)\, S \\
&= fR(X,Y)S.
\end{aligned}$$

iv) For the final part, we first recall from the definition of the commutator that

$$X(Yf) - Y(Xf) - [X,Y]f = 0 \tag{2.21}$$

for any sufficiently smooth function. We will apply this with $f = g(Z,W)$. Using the fact that ∇ is the Levi-Civita connection, we have:

$$\begin{aligned}
X\left(Y\left[g\left(Z,W\right)\right]\right) &= X\left(g\left(\nabla_Y Z, W\right) + g(Z,\nabla_Y W\right) \\
&= g\left(\nabla_X\nabla_Y Z, W\right) + g\left(\nabla_Y Z, \nabla_X W\right) \\
&\quad + g\left(\nabla_X Z, \nabla_Y W\right) + g\left(Z, \nabla_X\nabla_Y W\right)
\end{aligned} \tag{2.22}$$

Similarly, we have

$$\begin{aligned}
Y\left(X\left[g\left(Z,W\right)\right]\right) &= g\left(\nabla_Y\nabla_X Z, W\right) + g\left(\nabla_X Z, \nabla_Y W\right) \\
&\quad + g\left(\nabla_Y Z, \nabla_X W\right) + g\left(Z, \nabla_Y\nabla_X W\right)
\end{aligned} \tag{2.23}$$

and

$$[X,Y]\left(g(Z,W)\right) = g\left(\nabla_{[X,Y]}Z, W\right) + g\left(Z, \nabla_{[X,Y]}W\right) \tag{2.24}$$

Taking (2.22)−(2.23)−(2.24) and noting a cancellation between terms with one derivative falling on Z and one on W, we arrive at the result.

□

Corollary 2.5. *Suppose $X, Y, Z \in \mathfrak{X}_{k-1}(\mathcal{M})$, and let $\{e_\mu\}$ be a local basis, whose dual basis is $\{e^\mu\}$. Then if $X = X^\mu e_\mu$, $Y = Y^\mu e_\mu$, $Z = Z^\mu e_\mu$, we can write:*

$$R(X,Y)Z = R^\tau{}_{\sigma\mu\nu}X^\mu Y^\nu Z^\sigma e_\tau$$

Where $R^\mu{}_{\nu\sigma\tau} \in C^{r-2}(\mathcal{M};\mathbb{R})$ are the components of a $(1,3)-$tensor, called the Riemann tensor*, given by:*

$$R^\tau{}_{\sigma\mu\nu} = e^\tau\left(R(e_\mu, e_\nu)e_\sigma\right).$$

Proof. We simply use the linearity results established above to write

$$\begin{aligned} R(X,Y)Z &= R(X^\mu e_\mu, Y^\nu e_\nu)\,[Z^\mu e_\nu] \\ &= X^\mu Y^\nu Z^\sigma R(e_\mu, e_\nu) e_\sigma \\ &= X^\mu Y^\nu Z^\sigma e^\tau\,[R(e_\mu, e_\nu)e_\sigma]\, e_\tau \\ &= R^\tau{}_{\sigma\mu\nu} X^\mu Y^\nu Z^\sigma e_\tau. \end{aligned}$$

□

From parts *i*), *iv*) of Lemma 2.9, it is straightforward to show that the Riemann tensor has the following symmetries:

$$R^\tau{}_{\sigma\mu\nu} = -R^\tau{}_{\sigma\nu\mu}, \qquad R_{\tau\sigma\mu\nu} = -R_{\sigma\tau\mu\nu},$$

where we recall that indices are raised and lowered with the metric.

Exercise 3.6. a) Suppose $\omega \in \mathfrak{X}^*_{k-1}(\mathcal{M})$ and $X, Y, Z \in \mathfrak{X}_{k-1}(\mathcal{M})$. By considering (2.21) with $f = \omega[Z]$, and recalling that $X(\omega[Z]) = (\nabla_X \omega)[Z] + \omega[\nabla_X Z]$, show that:

$$\omega\,[R(X,Y)Z] + (R(X,Y)\omega)\,Z = 0.$$

b) Deduce that in a local basis, the action of $R(X,Y)$ on a one-form is given by:

$$R(X,Y)\omega = -R^\tau{}_{\sigma\mu\nu} X^\mu Y^\nu \omega_\tau e^\sigma$$

c) Show that if S_1, S_2 are $C^{k-1}-$regular tensor fields, then

$$R(X,Y)\,[S_1 \otimes S_2] = [R(X,Y)S_1] \otimes S_2 + S_1 \otimes [R(X,Y)S_2]\,.$$

d) Deduce that in a local basis, the action of $R(X,Y)$ on an arbitrary $(p,q)-$tensor is given by:

$$\begin{aligned}[R(X,Y)S]^{\mu_1\ldots\mu_p}{}_{\nu_1\ldots\nu_q} = \Big[& R^{\mu_1}{}_{\sigma\mu\nu} S^{\sigma\mu_2\ldots\mu_p}{}_{\nu_1\ldots\nu_q} + \cdots + R^{\mu_p}{}_{\sigma\mu\nu} S^{\mu_1\ldots\mu_{p-1}\sigma}{}_{\nu_1\ldots\nu_q} \\ & - R^\sigma{}_{\nu_1\mu\nu} S^{\mu_1\ldots\mu_p}{}_{\sigma\nu_2\ldots\nu_q} - \cdots - R^\sigma{}_{\nu_q\mu\nu} S^{\mu_1\ldots\mu_p}{}_{\nu_1\ldots\nu_{q-1}\sigma}\Big] X^\mu Y^\nu \end{aligned}$$

The exercise above shows that the Riemann tensor on its own is sufficient to determine the action of $R(X,Y)$ on an arbitrary tensor.

The Riemann tensor has further symmetries, which are encapsulated in the Bianchi identities. We will simply state these at this stage.

Theorem 2.6. *The following identities hold for any vector fields $X, Y, Z, W \in \mathfrak{X}_{k-1}(\mathcal{M})$:*

$$\begin{aligned} R(X,Y)Z + R(Y,Z)X + R(Z,X)Y &= 0 && \textit{1}^{st} \textit{ Bianchi identity} \\ [\nabla_W R]\,(X,Y)Z + [\nabla_X R]\,(Y,W)Z + [\nabla_Y R]\,(W,X)Z &= 0 && \textit{2}^{nd} \textit{ Bianchi identity} \end{aligned}$$

Proof. The proof is essentially by direct calculation, which can be simplified somewhat by choosing the vector fields cleverly. See for example the book of Do Carmo, “Riemannian Geometry”, pp 91, 106. □

Corollary 2.7. *With respect to a local basis, the first Bianchi identity can be written:*

$$R^{\tau}{}_{\sigma\mu\nu} + R^{\tau}{}_{\mu\nu\sigma} + R^{\tau}{}_{\nu\sigma\mu} = 0. \tag{2.25}$$

This, together with the previously established antisymmetry properties implies:

$$R_{\tau\sigma\mu\nu} = R_{\mu\nu\tau\sigma}$$

With this in mind, we can write the second Bianchi identity is equivalent to:

$$\nabla_\sigma R_{\mu\nu\tau\lambda} + \nabla_\mu R_{\nu\sigma\tau\lambda} + \nabla_\nu R_{\sigma\mu\tau\lambda} = 0. \tag{2.26}$$

Proof. Writing the 1^{st} Bianchi identity with respect to a local basis, we have:

$$R^{\tau}{}_{\sigma\mu\nu} X^\mu Y^\nu Z^\sigma e_\tau + R^{\tau}{}_{\sigma\mu\nu} Y^\mu Z^\nu X^\sigma e_\tau + R^{\tau}{}_{\sigma\mu\nu} Z^\mu X^\nu Y^\sigma e_\tau = 0$$

relabelling indices, and pulling out factors, this is equivalent to:

$$\left(R^{\tau}{}_{\sigma\mu\nu} + R^{\tau}{}_{\mu\nu\sigma} + R^{\tau}{}_{\nu\sigma\mu}\right) X^\mu Y^\nu Z^\sigma e_\tau = 0$$

which gives (2.25) since the vector fields X, Y, Z are arbitrary and $\{e_\mu\}$ is a basis.

Now, consider the same identity written out four times, with indices all lowered:

$$\begin{aligned}
0 &= R_{\tau\sigma\mu\nu} + R_{\tau\mu\nu\sigma} + R_{\tau\nu\sigma\mu} \\
0 &= R_{\sigma\mu\nu\tau} + R_{\sigma\nu\tau\mu} + R_{\sigma\tau\mu\nu} \\
0 &= R_{\mu\nu\tau\sigma} + R_{\mu\tau\sigma\nu} + R_{\mu\sigma\nu\tau} \\
0 &= R_{\nu\tau\sigma\mu} + R_{\nu\sigma\mu\tau} + R_{\nu\mu\tau\sigma}
\end{aligned}$$

Adding these four identities and using the antisymmetry on the first and last pairs of indices, all of the terms in the first column cancel against terms in the last column, and we have:

$$0 = 2R_{\tau\mu\nu\sigma} - 2R_{\nu\sigma\tau\mu}$$

which gives the result.

For the last part, we can write the second Bianchi identity with respect to the local basis as:

$$\left(\nabla_\sigma R^{\tau}{}_{\lambda\mu\nu}\right) W^\sigma X^\mu Y^\nu Z^\lambda e_\tau + \left(\nabla_\sigma R^{\tau}{}_{\lambda\mu\nu}\right) X^\sigma Y^\mu W^\nu Z^\lambda e_\tau + \left(\nabla_\sigma R^{\tau}{}_{\lambda\mu\nu}\right) Y^\sigma W^\mu X^\nu Z^\lambda e_\tau = 0$$

Relabelling, this becomes

$$\left(\nabla_\sigma R^{\tau}{}_{\lambda\mu\nu} + \nabla_\mu R^{\tau}{}_{\lambda\nu\sigma} + \nabla_\nu R^{\tau}{}_{\lambda\sigma\mu}\right) W^\sigma X^\mu Y^\nu Z^\lambda e_\tau = 0$$

Thus the second Bianchi identity is equivalent to the vanishing of the bracket above. Lowering τ and using the interchange symmetry we have:

$$\nabla_\sigma R_{\mu\nu\tau\lambda} + \nabla_\mu R_{\nu\sigma\tau\lambda} + \nabla_\nu R_{\sigma\mu\tau\lambda} = 0. \qquad \square$$

2.3.2 Ricci and scalar curvature

From the Riemann tensor it is possible to construct other curvature tensors, which capture some aspect or other of the geometry of the manifold. An important curvature is the Ricci curvature, which is a $C^{r-2}-$regular $(0,2)-$tensor, where for $X, Y \in T_p(\mathcal{M})$, $Ric_g(X,Y)|_p$ is defined as the trace of the endomorphism:

$$\mathscr{R}: \begin{array}{ccc} T_p\mathcal{M} & \to & T_p\mathcal{M} \\ Z & \mapsto & R(Z,X)Y \end{array}$$

In a local basis, we have

$$Ric_g(X,Y) = R^{\tau}{}_{\mu\tau\nu} X^{\mu} Y^{\nu} = R_{\mu\nu} X^{\mu} Y^{\nu},$$

where we define:

$$R_{\mu\nu} := Ric_g(e_{\mu}, e_{\nu}).$$

by the interchange symmetry of the Riemann tensor, the Ricci tensor is symmetric, $Ric_g(X,Y) = Ric_g(Y,X)$.

We also define the scalar curvature, sometimes called the Ricci scalar, R_g, which is the trace of Ric_g with respect to the metric. In local coordinates:

$$R_g = g^{\mu\nu} R_{\mu\nu}$$

Lemma 2.10. *The contracted Bianchi identities hold:*

$$\mathrm{div}_g \left(Ric_g - \frac{1}{2} R_g g \right) = 0.$$

Proof. We work in a local basis. Consider the second Bianchi identity in components (2.26):

$$\nabla_{\sigma} R_{\mu\nu\tau\lambda} + \nabla_{\mu} R_{\nu\sigma\tau\lambda} + \nabla_{\nu} R_{\sigma\mu\tau\lambda} = 0.$$

Contracting with $g^{\mu\tau}$, we have:

$$\nabla_{\sigma} R_{\nu\lambda} + \nabla^{\tau} R_{\nu\sigma\tau\lambda} - \nabla_{\nu} R_{\sigma\lambda} = 0.$$

Contracting again with $g^{\sigma\lambda}$, we obtain:

$$2\nabla^{\sigma} R_{\nu\sigma} - \nabla_{\nu} R_g = 0$$

Dividing by 2 and re-writing this, we have:

$$\nabla^{\sigma} \left(R_{\nu\sigma} - \frac{1}{2} R_g g_{\nu\sigma} \right) = 0,$$

which is the result. □

The tensor field $Ric_g - \frac{1}{2} R_g g$ is often referred to as the Einstein tensor, and denoted G.

Example 12. From Example 9 we have all of the information we require to calculate the Riemann tensor for the Schwarzschild metric in Painlevé-Gullstrand coordinates, with respect to the orthonormal basis $\{e_\mu\}$ given in Example 7. We calculate, for example:

$$
\begin{aligned}
R(e_0, e_1)e_0 &= \nabla_{e_0}\nabla_{e_1}e_0 - \nabla_{e_1}\nabla_{e_0}e_0 - \nabla_{[e_0,e_1]}e_0 \\
&= \nabla_{e_0}\left(\frac{1}{2r}\sqrt{\frac{2m}{r}}e_1\right) - \nabla_{e_1}(0) - \nabla_{-\frac{1}{2r}\sqrt{\frac{2m}{r}}e_1}e_0 \\
&= -\sqrt{\frac{2m}{r}}\frac{\partial}{\partial r}\left(\frac{1}{2r}\sqrt{\frac{2m}{r}}\right)e_1 + \frac{1}{2r}\sqrt{\frac{2m}{r}}\frac{1}{2r}\sqrt{\frac{2m}{r}}e_1 \\
&= \frac{2m}{r^3}e_1
\end{aligned}
$$

In this way, we can calculate:

$$
\begin{aligned}
&R(e_0, e_1)e_0 = \frac{2m}{r^3}e_1, && R(e_0, e_1)e_1 = \frac{2m}{r^3}e_0, && R(e_0, e_1)e_A = 0, \\
&R(e_0, e_A)e_0 = -\frac{m}{r^3}e_A, && R(e_0, e_A)e_1 = 0, && R(e_0, e_A)e_B = -\frac{m}{r^3}\delta_{AB}e_0, \\
&R(e_1, e_A)e_0 = 0, && R(e_1, e_A)e_1 = \frac{m}{r^3}e_A, && R(e_1, e_A)e_B = -\frac{m}{r^3}\delta_{AB}e_1, \\
&R(e_A, e_B)e_0 = 0, && R(e_A, e_B)e_1 = 0, && R(e_A, e_B)e_C = \frac{2m}{r^3}\left(\delta_{BC}e_A - \delta_{AC}e_B\right),
\end{aligned}
$$

Or in terms of components, we can extract the non-trivial components of the Riemann tensor:

$$
\begin{aligned}
R_{0101} &= -\frac{2m}{r^3}, & R_{0A0B} &= \frac{m}{r^3}\delta_{AB}, \\
R_{1A1B} &= -\frac{m}{r^3}\delta_{AB}, & R_{ABCD} &= \frac{2m}{r^3}\left(\delta_{BD}\delta_{AC} - \delta_{AD}\delta_{BC}\right),
\end{aligned}
$$

with all other components either related to these by symmetries of the Riemann tensor, or else vanishing.

To calculate the Ricci tensor, we have to take the trace over the first and third indices of the Riemann tensor. Since we work in an orthonormal basis, this is relatively straightforward, and we find, for example:

$$
\begin{aligned}
R_{00} = R^\mu{}_{0\mu 0} &= R_{1010} + \delta^{AB}R_{A0B0} \\
&= -\frac{2m}{r^3} + \frac{m}{r^3}\delta_{AB}\delta^{AB} = 0.
\end{aligned}
$$

Similarly:

$$
\begin{aligned}
R_{11} = R^\mu{}_{0\mu 0} &= -R_{0101} + \delta^{AB}R_{A1B1} \\
&= \frac{2m}{r^3} - \frac{m}{r^3}\delta_{AB}\delta^{AB} = 0.
\end{aligned}
$$

and finally:

$$\begin{aligned} R_{AB} = R^{\mu}{}_{A\mu B} &= -R_{0A0B} + R_{1A1B} + R_{CADB}\delta^{CD} \\ &= -\frac{m}{r^3}\delta_{AB} - \frac{m}{r^3}\delta_{AB} + \frac{2m}{r^3}\delta^{CD}\left(\delta_{CD}\delta_{AB} - \delta_{AD}\delta_{BC}\right) \\ &= 0. \end{aligned}$$

We thus conclude that the Ricci tensor (and hence also the scalar curvature) of the Schwarzschild metric vanishes identically!

Note that the full curvature does *not* vanish: in fact, we see that the curvature becomes singular near $r = 0$, which is the location of the black hole singularity.

2.3.3 Local expressions

We can work out an expression for the various curvatures in terms of the connection coefficients, and ultimately the components of the metric. To do this, we simply have to insert expressions for various derivatives.

Lemma 2.11. *Suppose $\{e_\mu\}$ is a local basis, with commutator coefficients given by:*

$$[e_\mu, e_\nu] = C^{\sigma}{}_{\mu\nu} e_\sigma,$$

and connection coefficients given by

$$\nabla_{e_\mu} e_\nu = \Gamma^{\sigma}{}_{\mu\nu} e_\sigma.$$

Then we have the following expression for the components of the Riemann tensor:

$$R^{\tau}{}_{\sigma\mu\nu} = e_\mu\left(\Gamma^{\tau}{}_{\nu\sigma}\right) + \Gamma^{\lambda}{}_{\nu\sigma}\Gamma^{\tau}{}_{\mu\lambda} - e_\nu\left(\Gamma^{\tau}{}_{\mu\sigma}\right) - \Gamma^{\lambda}{}_{\mu\sigma}\Gamma^{\tau}{}_{\nu\lambda} - \Gamma^{\tau}{}_{\lambda\sigma}C^{\lambda}{}_{\mu\nu} \tag{2.27}$$

For the Ricci tensor, we have:

$$R_{\sigma\nu} = e_\mu\left(\Gamma^{\mu}{}_{\nu\sigma}\right) + \Gamma^{\lambda}{}_{\nu\sigma}\Gamma^{\mu}{}_{\mu\lambda} - e_\nu\left(\Gamma^{\mu}{}_{\mu\sigma}\right) - \Gamma^{\lambda}{}_{\mu\sigma}\Gamma^{\mu}{}_{\nu\lambda} - \Gamma^{\mu}{}_{\lambda\sigma}C^{\lambda}{}_{\mu\nu}$$

Proof. We calculate

$$\begin{aligned} \nabla_{e_\mu}\nabla_{e_\nu} e_\sigma &= \nabla_{e_\mu}\left(\Gamma^{\lambda}{}_{\nu\sigma} e_\lambda\right) \\ &= e_\mu\left(\Gamma^{\lambda}{}_{\nu\sigma}\right) e_\lambda + \Gamma^{\lambda}{}_{\nu\sigma}\Gamma^{\tau}{}_{\mu\lambda} e_\tau \\ &= \left[e_\mu\left(\Gamma^{\tau}{}_{\nu\sigma}\right) + \Gamma^{\lambda}{}_{\nu\sigma}\Gamma^{\tau}{}_{\mu\lambda}\right] e_\tau \end{aligned}$$

Similarly,

$$\nabla_{e_\nu}\nabla_{e_\mu} e_\sigma = \left[e_\nu\left(\Gamma^{\tau}{}_{\mu\sigma}\right) + \Gamma^{\lambda}{}_{\mu\sigma}\Gamma^{\tau}{}_{\nu\lambda}\right] e_\tau.$$

We also have

$$\begin{aligned} \nabla_{[e_\mu, e_\nu]} e_\sigma &= \nabla_{C^{\lambda}{}_{\mu\nu} e_\lambda} e_\sigma \\ &= C^{\lambda}{}_{\mu\nu}\nabla_{e_\lambda} e_\sigma \\ &= \Gamma^{\mu}{}_{\lambda\sigma} C^{\lambda}{}_{\mu\nu} e_\tau \end{aligned}$$

Now, recalling that

$$R(e_\mu, e_\nu)e_\sigma = \nabla_{e_\mu}\nabla_{e_\nu}e_\sigma - \nabla_{e_\nu}\nabla_{e_\mu}e_\sigma - \nabla_{[e_\mu,e_\nu]}e_\sigma = R^\tau{}_{\sigma\mu\nu}e_\tau$$

we have the expression above for the Riemann tensor. The expression for the Ricci tensor follows from contracting on τ, μ. □

This expression allows us to calculate the form of the Ricci tensor, thought of as a (nonlinear) partial differential operator acting on g.

Corollary 2.8. *In a local coordinate basis, we can write:*

$$R_{\alpha\sigma\mu\nu} = \frac{1}{2}\left(\frac{\partial^2 g_{\alpha\nu}}{\partial x^\mu \partial x^\sigma} + \frac{\partial^2 g_{\sigma\mu}}{\partial x^\nu \partial x^\alpha} - \frac{\partial^2 g_{\alpha\mu}}{\partial x^\nu \partial x^\sigma} - \frac{\partial^2 g_{\sigma\nu}}{\partial x^\mu \partial x^\alpha}\right) - \Gamma_{\lambda\mu\alpha}\Gamma^\lambda{}_{\nu\sigma} + \Gamma_{\lambda\nu\alpha}\Gamma^\lambda{}_{\mu\sigma}. \tag{2.28}$$

and

$$\begin{aligned} R_{\sigma\nu} = &-\frac{1}{2}g^{\mu\alpha}\frac{\partial^2 g_{\sigma\nu}}{\partial x^\mu \partial x^\alpha} \\ &+ \frac{1}{2}\frac{\partial}{\partial x^\sigma}\Gamma_{\nu\mu}{}^\mu + \frac{1}{2}\frac{\partial}{\partial x^\nu}\Gamma_{\sigma\mu}{}^\mu - \Gamma_{\lambda\mu}{}^\mu\Gamma^\lambda{}_{\nu\sigma} \\ &+ \Gamma_{\tau\lambda\nu}\Gamma^{\tau\lambda}{}_\sigma + \Gamma_{\tau\lambda\nu}\Gamma_\sigma{}^{\tau\lambda} + \Gamma_{\tau\lambda\sigma}\Gamma_\nu{}^{\tau\lambda} \end{aligned} \tag{2.29}$$

Proof. 1. As a short hand, we will use $\partial_\alpha = \frac{\partial}{\partial x^\alpha}$, $\partial^2_{\alpha\beta} = \frac{\partial^2}{\partial x^\alpha \partial x^\beta}$. We have to be careful with the partial derivatives, since unlike covariant derivatives they do *not* commute with the raising and lowering of indices. Now recall:

$$\Gamma_{\alpha\beta\tau} = g_{\alpha\tau}\Gamma^\tau{}_{\beta\tau} = \frac{1}{2}\left(\partial_\beta g_{\alpha\tau} + \partial_\tau g_{\alpha\beta} - \partial_\alpha g_{\beta\tau}\right)$$

so that

$$\partial_\beta g_{\alpha\tau} = \Gamma_{\alpha\beta\tau} + \Gamma_{\tau\beta\alpha} \tag{2.30}$$

As a consequence, we have

$$g_{\alpha\tau}\partial_\beta\left(\Gamma^\tau{}_{\mu\nu}\right) = \partial_\beta\left(\Gamma_{\alpha\mu\nu}\right) - \left(\Gamma_{\alpha\beta\tau} + \Gamma_{\tau\beta\alpha}\right)\Gamma^\tau{}_{\mu\nu}.$$

2. Inserting this into the expression (2.27) and recalling that in a coordinate basis we have $C^\lambda{}_{\mu\nu} \equiv 0$, we find:

$$\begin{aligned} R_{\alpha\sigma\mu\nu} &= g_{\alpha\tau}R^\tau{}_{\sigma\mu\nu} \\ &= g_{\alpha\tau}\partial_\mu\left(\Gamma^\tau{}_{\nu\sigma}\right) - g_{\alpha\tau}\partial_\nu\left(\Gamma^\tau{}_{\mu\sigma}\right) + \Gamma^\lambda{}_{\nu\sigma}\Gamma_{\alpha\mu\lambda} - \Gamma^\lambda{}_{\mu\sigma}\Gamma_{\alpha\nu\lambda} \\ &= \partial_\mu\left(\Gamma_{\alpha\nu\sigma}\right) - \left(\Gamma_{\alpha\mu\tau} + \Gamma_{\tau\mu\alpha}\right)\Gamma^\tau{}_{\nu\sigma} \\ &\quad - \partial_\nu\left(\Gamma_{\alpha\mu\sigma}\right) + \left(\Gamma_{\alpha\nu\tau} + \Gamma_{\tau\nu\alpha}\right)\Gamma^\tau{}_{\mu\sigma} \\ &\quad + \Gamma^\lambda{}_{\nu\sigma}\Gamma_{\alpha\mu\lambda} - \Gamma^\lambda{}_{\mu\sigma}\Gamma_{\alpha\nu\lambda} \\ &= \partial_\mu\left(\Gamma_{\alpha\nu\sigma}\right) - \partial_\nu\left(\Gamma_{\alpha\mu\sigma}\right) - \Gamma_{\tau\mu\alpha}\Gamma^\tau{}_{\nu\sigma} + \Gamma_{\tau\nu\alpha}\Gamma^\tau{}_{\mu\sigma} \end{aligned}$$

Inserting the definition of Γ into the first two terms, we calculate

$$\begin{aligned}R_{\alpha\sigma\mu\nu} &= \frac{1}{2}\left(\partial^2_{\mu\nu}g_{\alpha\sigma} + \partial^2_{\mu\sigma}g_{\alpha\nu} - \partial^2_{\mu\alpha}g_{\nu\sigma} - \partial^2_{\nu\mu}g_{\alpha\sigma} - \partial^2_{\nu\sigma}g_{\alpha\mu} + \partial^2_{\nu\alpha}g_{\mu\sigma}\right) \\ &\quad - \Gamma_{\tau\mu\alpha}\Gamma^{\tau}{}_{\nu\sigma} + \Gamma_{\tau\nu\alpha}\Gamma^{\tau}{}_{\mu\sigma} \\ &= \frac{1}{2}\left(\partial^2_{\mu\sigma}g_{\alpha\nu} + \partial^2_{\nu\alpha}g_{\mu\sigma} - \partial^2_{\mu\alpha}g_{\nu\sigma} - \partial^2_{\nu\sigma}g_{\alpha\mu}\right) - \Gamma_{\tau\mu\alpha}\Gamma^{\tau}{}_{\nu\sigma} + \Gamma_{\tau\nu\alpha}\Gamma^{\tau}{}_{\mu\sigma}\end{aligned}$$

which is (2.28).

3. To find the Ricci tensor, we first differentiate (2.30) to find:

$$\begin{aligned}\partial^2_{\mu\sigma}g_{\alpha\nu} &= \partial_\sigma\left(\Gamma_{\alpha\mu\nu} + \Gamma_{\nu\mu\alpha}\right) \\ \partial^2_{\nu\alpha}g_{\mu\sigma} &= \partial_\nu\left(\Gamma_{\mu\alpha\sigma} + \Gamma_{\sigma\alpha\mu}\right) \\ \partial^2_{\nu\sigma}g_{\alpha\mu} &= \frac{1}{2}\partial_\nu\left(\Gamma_{\alpha\sigma\mu} + \Gamma_{\mu\sigma\alpha}\right) + \frac{1}{2}\partial_\sigma\left(\Gamma_{\alpha\nu\mu} + \Gamma_{\mu\nu\alpha}\right)\end{aligned}$$

taking the sum of the first two equalities and subtracting the third, we have

$$\begin{aligned}\partial^2_{\mu\sigma}g_{\alpha\nu} + \partial^2_{\nu\alpha}g_{\mu\sigma} - \partial^2_{\nu\sigma}g_{\alpha\mu} &= \partial_\sigma\left(\Gamma_{\nu\mu\alpha}\right) + \partial_\nu\left(\Gamma_{\sigma\mu\alpha}\right) \\ &\quad + \frac{1}{2}\left[\partial_\sigma\left(\Gamma_{\alpha\mu\nu}\right) - \partial_\sigma\left(\Gamma_{\mu\alpha\nu}\right) + \partial_\nu\left(\Gamma_{\alpha\mu\sigma}\right) - \partial_\nu\left(\Gamma_{\mu\alpha\sigma}\right)\right]\end{aligned}$$

Notice that the term in square brackets is antisymmetric in α and μ, so that

$$\begin{aligned}g^{\alpha\mu}\left(\partial^2_{\mu\sigma}g_{\alpha\nu} + \partial^2_{\nu\alpha}g_{\mu\sigma} - \partial^2_{\nu\sigma}g_{\alpha\mu}\right) &= g^{\mu\alpha}\left(\partial_\sigma\left(\Gamma_{\nu\mu\alpha}\right) + \partial_\nu\left(\Gamma_{\sigma\mu\alpha}\right)\right) \\ &= \partial_\sigma\left(\Gamma_{\nu\mu}{}^{\mu}\right) + \partial_\nu\left(\Gamma_{\sigma\mu}{}^{\mu}\right) \\ &\quad - \Gamma_{\nu\mu\alpha}\partial_\sigma g^{\mu\alpha} - \Gamma_{\sigma\mu\alpha}\partial_\nu g^{\mu\alpha}\end{aligned}$$

Now, since $\partial_\mu g^{\alpha\beta} = -g^{\lambda\alpha}\left(\partial_\sigma g_{\lambda\tau}\right)g^{\beta\tau}$, we have:

$$\begin{aligned}-\Gamma_{\nu\mu\alpha}\partial_\sigma g^{\mu\alpha} - \Gamma_{\sigma\mu\alpha}\partial_\nu g^{\mu\alpha} &= \Gamma_\nu{}^{\mu\alpha}\partial_\sigma g_{\mu\alpha} + \Gamma_\sigma{}^{\mu\alpha}\partial_\nu g_{\mu\alpha} \\ &= \Gamma_\nu{}^{\mu\alpha}\left(\Gamma_{\mu\sigma\alpha} + \Gamma_{\alpha\sigma\mu}\right) + \Gamma_\sigma{}^{\mu\alpha}\left(\Gamma_{\mu\nu\alpha} + \Gamma_{\alpha\nu\mu}\right) \\ &= 2\Gamma_{\tau\lambda\nu}\Gamma_\sigma{}^{\tau\lambda} + 2\Gamma_{\tau\lambda\sigma}\Gamma_\nu{}^{\tau\lambda}\end{aligned}$$

4. To obtain (2.29) we multiply the expression (2.28) by $g^{\alpha\mu}$ and use the result of part 3. above to rewrite three of the four $\partial^2 g$ terms in terms of Γ

□

The form that we have put the Ricci tensor in might seem a bit strange: we've made some second derivatives of the metric explicit and others we have written in terms of the connection components. To explain why we have done this, at this stage it's useful to introduce a particular choice of coordinates, known as *wave coordinates*. Recall that the wave equation is given in a local coordinate basis by (2.14):

$$0 = \Box_g f = g^{\mu\nu}\frac{\partial^2 f}{\partial x^\mu \partial x^\nu} - g^{\mu\nu}\Gamma^\sigma{}_{\mu\nu}\frac{\partial f}{\partial x^\sigma}.$$

We say that we are working in wave coordinates if the coordinate functions x^α are themselves solutions of the wave equation, so that $\Box_g x^\alpha = 0$. Since $\partial_\sigma x^\alpha = \delta^\alpha{}_\sigma$, we get the following condition on the connection components when expressed in a local basis induced by wave coordinates:

$$0 = -g^{\mu\nu}\Gamma^\alpha{}_{\mu\nu} = \Gamma^\alpha{}_\mu{}^\mu. \tag{2.31}$$

Theorem 2.9. *With respect to wave coordinates, the Ricci tensor takes the form:*

$$R_{\sigma\nu} = R^{(w)}_{\sigma\nu} := -\frac{1}{2}g^{\mu\alpha}\frac{\partial^2 g_{\sigma\nu}}{\partial x^\mu \partial x^\alpha} + P_{\sigma\nu}(g, \partial g) \tag{2.32}$$

Here P is a homogeneous quadratic form in the first derivatives,

$$P_{\sigma\nu}(g, \partial g) = P^{\alpha\beta\gamma,\lambda\tau\mu}_{\sigma\nu}(g)\partial_\alpha g_{\beta\gamma}\partial_\lambda g_{\tau\mu},$$

where $P^{\alpha\beta\gamma,\lambda\tau\mu}_{\sigma\nu}(g) = P^{\lambda\tau\mu,\alpha\beta\gamma}_{\sigma\nu}(g)$ is determined from:

$$P^{\alpha\beta\gamma,\lambda\tau\mu}_{\sigma\nu}(g)\partial_\alpha g_{\beta\gamma}\partial_\lambda g_{\tau\mu} = \Gamma_{\tau\lambda\nu}\Gamma^{\tau\lambda}{}_\sigma + \Gamma_{\tau\lambda\nu}\Gamma_\sigma{}^{\tau\lambda} + \Gamma_{\tau\lambda\sigma}\Gamma_\nu{}^{\tau\lambda}. \tag{2.33}$$

Proof. We simply insert the condition $0 = \Gamma^\alpha{}_\mu{}^\mu$ into (2.29), which implies that all of the terms on the second line vanish. The fact that P is a homogeneous quadratic form in the first derivatives follows from the fact that $\Gamma_{\alpha\beta\gamma}$ is linear in the first derivatives, and that the right hand side of (2.33) is a symmetric quadratic form in $\Gamma_{\alpha\beta\gamma}$. $\Box$

This shows that the Ricci tensor (in appropriate coordinates) can be though of as a quasilinear wave operator acting on the metric tensor.

Chapter 3

Einstein's equations

3.1 Einstein's equations and matter models

We are now in a position to write down Einstein's equations for the gravitational field, which describe a 4−dimensional Lorentzian manifold $(\mathcal{M}, g)$:

$$Ric_g - \frac{1}{2}Rg + \Lambda g = T. \tag{3.1}$$

The left hand side of Einstein's equations involves terms constructed from the metric g. The first two terms are familiar from the discussion of the previous chapter. The third term on the left hand side is a constant multiple of the metric itself. Λ here is a parameter of the theory known as the *cosmological constant*. It was introduced by Einstein in order that the theory would admit solutions corresponding to a stationary universe. The discovery by Hubble that in fact galaxies are all moving apart caused Einstein to throw away this term, dismissing it as the "greatest blunder" of his life. Modern measurements of the Cosmic Microwave Background, and Type Ia Supernovae data suggest that the cosmological constant term should be present, and that $\Lambda > 0$.

The term on the right hand side, T, is the energy-momentum tensor of the matter present in the spacetime. It is a symmetric, divergence free tensor. In order to close the system of equations represented by (3.1), we have to specify some model for the matter present in the spacetime, which describes how the matter evolves in time. Possible matter models include:

1. **Vacuum**. For this we set $T \equiv 0$, so that there is no matter present in the spacetime. Bothe the Minkowski spacetime, and the Schwarzschild spacetime that we have already encountered are solutions of the vacuum Einstein equations with $\Lambda = 0$.

2. **Wave matter** The matter content is encoded in a single function ψ satisfying the wave equation:

$$\Box_g \psi = 0$$

where $\Box_g$ is the wave operator of the metric g. The energy-momentum tensor is then given in a local basis by:

$$T_{\mu\nu}[\psi] = \nabla_\mu \psi \nabla_\nu \psi - \frac{1}{2} g_{\mu\nu} \nabla_\sigma \psi \nabla^\sigma \psi.$$

3. **Electromagnetic**. Here the matter content is encoded in an antisymmetric $(0,2)-$tensor, F satisfying the Maxwell equations, which in a local basis take the form:
$$\nabla_\mu F^\mu{}_\nu = 0, \qquad \nabla_{[\mu} F_{\nu\sigma]} = 0,$$
where ∇ is the Levi-Civita connection of g. The energy-momentum tensor is then given in a local basis by:
$$T_{\mu\nu}[F] = F_{\mu\sigma} F_\nu{}^\sigma - \frac{1}{4}\eta_{\mu\nu} F_{\sigma\tau} F^{\sigma\tau}.$$

4. **Perfect fluid** A perfect fluid is described by a local velocity $U \in \mathfrak{X}(\mathcal{M})$, which is everywhere a unit timelike vector field, together with a pressure p and a density ρ. They satisfy the first law of thermodynamics:
$$U[\rho] + (\rho + p)\mathrm{div}_g U = 0,$$
and Euler's equation:
$$(\rho + p)\nabla_U U + \mathrm{grad}_g p + U[p]U = 0,$$
This gives a closed system once a relation, called an equation of state, $p = p(\rho)$ is specified. The energy momentum tensor is then given by
$$T_{\mu\nu} = (\rho + p)U_\mu U_\nu + p g_{\mu\nu}.$$

All of these matter models are used for various purposes in the study of relativity. Notice that in general, the equations of motion for the matter fields depend on the metric, and of course the metric evolves according to Einstein's equations. This was summed up by Wheeler as:

> "Space acts on matter, telling it how to move. In turn, matter reacts back on space, telling it how to curve."[1]

In general, we have a complicated system of nonlinear, hyperbolic PDEs for the 10 components of the metric and the matter fields. For most of the rest of the course, we will focus on the vacuum case. This allows us to consider some of the challenges of studying general relativity in a somewhat simpler setting.

3.2 The linearised Einstein equations

A starting point for the study of any nonlinear PDE is often to study the linearisation about a known solution. This usually results in a simpler problem, which can be attacked with standard methods. The knowledge one gains from studying the linearised problem can then be used to try and tackle the full, nonlinear, problem. We will consider the

[1] C. W. Misner, K. S. Thorne and J. A. Wheeler, *Gravitation*. W. H. Freeman, 1973.

problem of linearising the vacuum Einstein equations about the Minkowski space. In the vacuum case, with $\Lambda = 0$, the equations reduce to

$$Ric_g = 0.$$

Recall that Minkowski space is the manifold $\mathbb{R}^4$, with coordinates $(x^\mu)_{\mu=0,\ldots 3}$ and the metric given by:

$$\eta = \eta_{\mu\nu} dx^\mu dx^\nu = -(dx^0)^2 + (dx^1)^2 + (dx^2)^2 + (dx^3)^2.$$

Now, since

$$\Box_\eta f = \eta^{\mu\nu} \frac{\partial^2 f}{\partial x^\mu \partial x^\nu},$$

the coordinates x^μ are wave coordinates for the Minkowski metric.

Let us suppose that we have a family of metrics ${}^{(s)}g$ defined on $\mathbb{R}^4$, with ${}^{(0)}g = \eta$ and which depend smoothly[2] on $s \in (-\epsilon, \epsilon)$. Suppose also that ${}^{(s)}g$ solves the Einstein equations, and that the coordinates x^μ are wave coordinates for ${}^{(s)}g$ for every $s \in (-\epsilon, \epsilon)$. This implies that

$$-\frac{1}{2} {}^{(s)}g^{\mu\alpha} \frac{\partial^2}{\partial x^\mu \partial x^\alpha} \left[{}^{(s)}g_{\sigma\nu} \right] + {}^{(s)}\Gamma_{\tau\lambda\nu} {}^{(s)}\Gamma^{\tau\lambda}{}_\sigma + {}^{(s)}\Gamma_{\tau\lambda\nu} {}^{(s)}\Gamma_\sigma{}^{\tau\lambda} + {}^{(s)}\Gamma_{\tau\lambda\sigma} {}^{(s)}\Gamma_\nu{}^{\tau\lambda} = 0$$

and

$$0 = {}^{(s)}\Gamma^{\alpha\mu}{}_\mu = {}^{(s)}g^{\alpha\tau} {}^{(s)}g^{\mu\nu} \left(\partial_\nu {}^{(s)}g_{\mu\tau} - \frac{1}{2} \partial_\tau {}^{(s)}g_{\mu\nu} \right).$$

To find the linearised Einstein equations, we differentiate these with respect to s, and then set $s = 0$. Recalling that ${}^{(0)}g_{\mu\nu} = \eta_{\mu\nu}$ and ${}^{(0)}\Gamma_{\tau\lambda\sigma} = 0$, we deduce that

$$0 = \Box_\eta \gamma_{\mu\nu}, \tag{3.2}$$

$$0 = \partial_\mu \gamma^\mu{}_\nu - \frac{1}{2} \partial_\nu \gamma_\sigma{}^\sigma, \tag{3.3}$$

where $\gamma_{\mu\nu} = \frac{d}{ds} \left[{}^{(s)}g_{\mu\nu} \right] \big|_{s=0}$, and indices are raised and lowered with η. Equation (3.2) simply says that the components of γ with respect to the canonical coordinates on Minkowski space each separately obey the wave equation. By the results of Chapter 1, a unique solution for $\gamma_{\mu\nu} \in C^\infty(\mathbb{R}^4)$ exists, provided we specify smooth initial data: $\gamma_{\mu\nu}|_{x^0=0}$ and $\partial_0 \gamma_{\mu\nu}|_{x^0=0}$. Can we simultaneously satisfy equation (3.3)? This represents a set of *constraints* that our solutions to (3.2) must satisfy. In order that the pair of equations (3.2), (3.3) admit any solutions at all, they must be compatible. That they are is a result of the following:

Lemma 3.1. *Suppose $\gamma_{\mu\nu}$ is a smooth solution of (3.2). Then (3.3) holds in $\mathbb{R}^4$ if, and only if:*

$$0 = \partial_\mu \gamma^\mu{}_\nu - \frac{1}{2} \partial_\nu \gamma_\sigma{}^\sigma \Big|_{x^0=0}, \tag{3.4}$$

$$0 = \partial_0 \left(\partial_\mu \gamma^\mu{}_\nu - \frac{1}{2} \partial_\nu \gamma_\sigma{}^\sigma \right) \Big|_{x^0=0}. \tag{3.5}$$

[2] In the sense that the components of ${}^{(s)}g$, and an appropriate number of their derivatives, with respect to any coordinate chart are smooth functions of s. The fact that such families of solutions exist is not a priori obvious, but happens to be true.

Proof. Let $F_\nu = \partial_\mu \gamma^\mu{}_\nu - \frac{1}{2}\partial_\nu \gamma_\sigma{}^\sigma$. Equation (3.3) is equivalent to $F_\nu = 0$. Notice that F_ν solves the wave equation for each ν:

$$\begin{aligned}\Box_\eta F_\nu &= \partial_\tau \partial^\tau \left(\partial_\mu \gamma^\mu{}_\nu - \frac{1}{2}\partial_\nu \gamma_\sigma{}^\sigma \right) \\ &= \partial_\mu \left(\Box_\eta \gamma^\mu{}_\nu \right) - \frac{1}{2}\partial_\nu \left(\Box_\eta \gamma_\sigma{}^\sigma \right) \\ &= 0.\end{aligned}$$

Now, by the uniqueness results of Chapter 1, we know that F_ν is uniquely determined by the value of F_ν, $\partial_0 F_\nu$ on the hyperplane $\{x^0 = 0\}$. In particular if $F_\nu = \partial_0 F_\nu = 0$ on $\{x^0 = 0\}$, then $F_\nu = 0$ in $\mathbb{R}^4$. □

What this result tells us is that it's enough to make sure that the constraint equations are satisfied at the initial time $x^0 = 0$, and the evolution equation (3.2) then ensures that the constraints *propagate* in time. The conditions (3.4), (3.5) will not hold for arbitrary choices of initial conditions $\gamma_{\mu\nu}|_{x^0=0}$ and $\partial_0 \gamma_{\mu\nu}|_{x^0=0}$, we need to restrict our choice of data to ensure that the initial constraints are satisfied.

Let us suppose that we are given $\phi, \beta_i, h_{ij}, k_{ij}$, for $i, j = 1, 2, 3$, which we assume to be smooth functions on $\mathbb{R}^3$. We suppose that the initial data for (3.2) is constructed from these functions in the following fashion:

$$\begin{aligned}\gamma_{00}|_{x^0=0} &= \phi \\ \gamma_{0i}|_{x^0=0} = \gamma_{i0}|_{x^0=0} &= \beta_i \\ \gamma_{ij}|_{x^0=0} &= h_{ij} \\ \partial_0 \gamma_{00}|_{x^0=0} &= 2k_{ii} \\ \partial_0 \gamma_{0j}|_{x^0=0} &= \partial_i h_{ij} - \frac{1}{2}\partial_j h_{ii} + \frac{1}{2}\partial_j \phi \\ \partial_0 \gamma_{ij}|_{x^0=0} &= -2k_{ij} + 2\partial_{(i}\beta_{j)}\end{aligned} \tag{3.6}$$

Lemma 3.2. *The solution $\gamma_{\mu\nu}$ to (3.2) with initial conditions (3.6) satisfies (3.3) throughout $\mathbb{R}^4$, and hence is a solution of the linearised Einstein equations, if and only if the following constraints hold on h_{ij}, k_{ij}:*

$$0 = \partial_i \partial_j h_{ij} - \partial_i \partial_i h_{jj}, \tag{3.7}$$

$$0 = \partial_i k_{ij} - \partial_j k_{ii}. \tag{3.8}$$

Proof. We first verify that constraint equation (3.4) is satisfied by our choice of initial data. Splitting into the time and space components, we first calculate:

$$\begin{aligned}\left.\partial_\mu \gamma^\mu{}_0 - \frac{1}{2}\partial_0 \gamma_\sigma{}^\sigma \right|_{x^0=0} &= \left. -\partial_0 \gamma_{00} + \partial_i \gamma_{i0} + \frac{1}{2}\partial_0 \gamma_{00} - \frac{1}{2}\partial_0 \gamma_{ii} \right|_{x^0=0} \\ &= -2k_{ii} + \partial_i \beta_i + k_{ii} - \frac{1}{2}\left(-2k_{ii} + 2\partial_i \beta_i\right) \\ &= 0.\end{aligned}$$

For the spacelike components we have:

$$\begin{aligned}
\partial_\mu \gamma^\mu{}_j - \frac{1}{2}\partial_j \gamma_\sigma{}^\sigma \Big|_{x^0=0} &= -\partial_0\gamma_{0j} + \partial_i\gamma_{ij} + \frac{1}{2}\partial_j\gamma_{00} - \frac{1}{2}\partial_j\gamma_{ii}\Big|_{x^0=0} \\
&= -\left.\partial_0\gamma_{0j}\right|_{x^0=0} + \partial_i h_{ij} + \frac{1}{2}\partial_j\phi - \frac{1}{2}\partial_j h_{ii} \\
&= 0.
\end{aligned}$$

Now we have to verify that (3.5) holds, at which point we are done by the previous Lemma. When we differentiate the constraint in the time direction, we will observe some components with two x^0-derivatives acting on them. To handle these, we use the fact that $\Box_\eta \gamma_{\mu\nu} = 0$, so that in particular:

$$\left.\partial_0\partial_0\gamma_{\mu\nu}\right|_{x^0=0} = \left.\partial_j\partial_j\gamma_{\mu\nu}\right|_{x^0=0}$$

We find that the $0-$component of (3.5) gives:

$$\begin{aligned}
\partial_0\left(\partial_\mu\gamma^\mu{}_0 - \frac{1}{2}\partial_0\gamma_\sigma{}^\sigma\right)\Big|_{x^0=0} &= -\partial_0\partial_0\gamma_{00} + \partial_0\partial_i\gamma_{i0} + \frac{1}{2}\partial_0\partial_0\gamma_{00} - \frac{1}{2}\partial_0\partial_0\gamma_{ii}\Big|_{x^0=0} \\
&= -\frac{1}{2}\partial_i\partial_i\phi + \partial_j\left(\partial_i h_{ij} - \frac{1}{2}\partial_j h_{ii} + \frac{1}{2}\partial_j\phi\right) - \frac{1}{2}\partial_j\partial_j h_{ii} \\
&= \partial_i\partial_j h_{ij} - \partial_i\partial_i h_{jj} \\
&= 0.
\end{aligned}$$

Where we use (3.7) in the last line. Finally, to verify the spacelike components of (3.5), we calculate:

$$\begin{aligned}
\partial_0\left(\partial_\mu\gamma^\mu{}_j - \frac{1}{2}\partial_j\gamma_\sigma{}^\sigma\right)\Big|_{x^0=0} &= -\partial_0\partial_0\gamma_{0j} + \partial_0\partial_i\gamma_{ij} + \frac{1}{2}\partial_j\partial_0\gamma_{00} - \frac{1}{2}\partial_j\partial_0\gamma_{ii}\Big|_{x^0=0} \\
&= -\partial_i\partial_i\beta_j + \partial_i\left(-2k_{ij} + \partial_i\beta_j + \partial_j\beta_i\right) \\
&\qquad + \partial_j k_{ii} - \frac{1}{2}\partial_j\left(-2k_{ii} + 2\partial_i\beta_i\right) \\
&= -2\left(\partial_i k_{ij} - \partial_j k_{ii}\right) \\
&= 0.
\end{aligned}$$

□

We can thus break the initial data down into geometrical objects defined on $\mathbb{R}^3$. We have two symmetric tensors, h and k, which have to obey the constraint equations (3.7), (3.8). We also have a scalar ϕ and a vector field β which are freely specifiable. Once we have specified these objects, there exists a unique solution γ to the equations (3.2) (3.3). We shall see later that h, k are *intrinsic* to the initial hypersurface $\{x^0 = 0\}$, while ϕ, β essentially encode information about the choice of coordinates (the wave coordinate condition doesn't fix completely fix the coordinates).

3.3 Hypersurface geometry and the constraint equations

[Before reading this section, you should make sure that you're familiar with the material in §A.3.4]

A note of caution We will specialise in this section to embedded spacelike submanifolds of Lorentzian manifolds. Many of the results have analogues in the case of embedded submanifolds of Riemannian manifolds. However, there are a few differences in sign introduced by the signature, so one should not assume that the formulae given here are directly valid in that situation.

3.3.1 The induced metric and second fundamental form

Let us suppose that we have a smooth, time oriented, four dimensional Lorentzian manifold $(\mathcal{M}, g)$. Suppose that we are also given a three dimensional manifold Σ. An *embedding* of Σ into $\mathcal{M}$ is a smooth map $\imath \in C^\infty(\Sigma; \mathcal{M})$, such that $\imath$ is a homeomorphism of Σ onto $\imath(\Sigma)$ and $\imath$ is an immersion, i.e. the push forward map $\imath_*$ acting on vectors is everywhere injective. As a result of the injectivity of $\imath_*$, we can identify $T_p\Sigma$ with a three-dimensional subspace $T_{\imath(p)}\imath(\Sigma) \subset T_{\imath(p)}\mathcal{M}$. We say that a vector $X \in T_{\imath(p)}\imath(\Sigma)$ is *tangent to* $\imath(\Sigma)$ *at* $\imath(p)$.

We assume now that an embedding has been fixed.

Definition 17. The metric induced on Σ by g is the pull-back of g to Σ by the embedding map $\imath$, and we denote the induced metric by $h := \imath^* g$. More concretely, for $X, Y \in T_p\Sigma$, we define:

$$h(X, Y) = g(\imath_* X, \imath_* Y)$$

We can translate our definitions of timelike/spacelike/null surfaces to the following:

Lemma 3.3. *The surface* $\imath(\Sigma)$ *is:*

i) Timelike *at* $\imath(p)$ *if and only if* h *is a Lorentzian metric at* p.

ii) Null *at* $\imath(p)$ *if and only if* h *is a degenerate quadratic form at* p.

iii) Spacelike *at* $\imath(p)$ *if and only if* h *is a Riemannian metric at* p.

We will mostly focus on the spacelike case, as this is the correct setting for an initial data surface for Einstein's equations. For each $p \in \Sigma$, there is a unique $N \in T_{\imath(p)}\mathcal{M}$ which is timelike, future directed, of unit length, and orthogonal to $T_{\imath(p)}\imath(\Sigma)$. Using the Canonical Immersion Theorem, Lemma A.8 in §A.3.4, we can assume that N is the restriction to $\imath(\Sigma)$ of a smooth vector field defined on $\mathcal{M}$. For any $V \in T_{\imath(p)}\mathcal{M}$, we define:

$$\top V := V + g(N, V)N, \qquad \perp V := -g(N, V)N$$

so that

$$V = \top V + \perp V$$

and we have $\top V \in T_{\imath(p)}\imath(\Sigma)$, and $\perp V$ is orthogonal to $T_{\imath(p)}\imath(\Sigma)$. In other words, $T_{\imath(p)}\mathcal{M}$ splits into

$$T_{\imath(p)}\mathcal{M} = T_{\imath(p)}\imath(\Sigma) \oplus N_{\imath(p)}\imath(\Sigma),$$

where $N_{\imath(p)}\imath(\Sigma) = \left(T_{\imath(p)}\imath(\Sigma)\right)^\perp$ is the orthogonal complement of $T_{\imath(p)}\imath(\Sigma)$ with respect to g. This is a one-dimensional timelike subspace, representing the normal directions to $\imath(\Sigma)$ with respect to the metric g.

We want to look at how the Levi-Civita connection ∇ behaves under this splitting. For this it will be useful to have the following result:

Lemma 3.4. *Let $\mathcal{U} \subset \Sigma$ be open, and suppose that $\imath : \Sigma \hookrightarrow \mathcal{M}$ is an embedding such that the image is spacelike.*

i) Suppose that $V_1, V_2, W \in \mathfrak{X}(\mathcal{M})$ satisfy $V_1 = V_2$ on $\imath(\mathcal{U})$. Then

$$\nabla_{V_1} W = \nabla_{V_2} W, \qquad \text{on} \imath(\mathcal{U}).$$

ii) Suppose that $W, V_1, V_2 \in \mathfrak{X}(\mathcal{M})$ satisfy $W \in T_{\imath(p)}\imath(\mathcal{U})$ for each $p \in \mathcal{U}$ and $V_1 = V_2$ on $\imath(\mathcal{U})$. Then

$$\nabla_W V_1 = \nabla_W V_2 \qquad \text{on} \imath(\mathcal{U}).$$

Proof. i) We have $V_1 - V_2 = 0$ on $\imath(\mathcal{U})$, so by the fact that ∇ is tensorial in its first slot, we have that on $\imath(\mathcal{U})$:

$$0 = \nabla_{V_1 - V_2} W = \nabla_{V_1} W - \nabla_{V_2} W$$

ii) It suffices to prove that if $V \in \mathfrak{X}(\mathcal{M})$ vanishes on $\imath(\mathcal{U})$, then $\nabla_W V = 0$ on $\imath(\mathcal{U})$. Let us fix some $K \in \mathfrak{X}(\mathcal{M})$ and define $f = g(V, K)|_{\imath(\mathcal{U})}$. Clearly $\imath^* f = 0$. Note also that there exists a vector field $X \in \mathfrak{X}(\mathcal{U})$ such that $\imath_* X = W$ on $\imath(\mathcal{U})$. We calculate that at $p \in \mathcal{U}$:

$$\begin{aligned} 0 &= X(\imath^* f)|_p \\ &= \imath_* X(f)|_{\imath(p)} = W(f)|_{\imath(p)} \\ &= W\left[g\left(V, K\right)\right]|_{\imath(p)} \\ &= g\left(\nabla_W V, K\right)|_{\imath(p)} + g\left(V, \nabla_W K\right)|_{\imath(p)} \\ &= g\left(\nabla_W V, K\right)|_{\imath(p)} \end{aligned}$$

Now, since K was arbitrary, we deduce $\nabla_W V|_{\imath(p)} = 0$. □

Suppose we have vector fields $X, Y \in \mathfrak{X}(\Sigma)$. By Corollary A.4, about any $p \in \Sigma$ we can find a neighbourhood $\mathcal{U}$ and two vector fields $\widetilde{X}, \widetilde{Y} \in \mathfrak{X}(\mathcal{M})$ such that $\imath_* X = \widetilde{X}$ and $\imath_* Y = \widetilde{Y}$ on $\imath(\mathcal{U})$, i.e. such that $\widetilde{X}, \widetilde{Y}$ extend X, Y away from $\imath(\mathcal{U})$. The previous Lemma shows that $\nabla_{\widetilde{X}} \widetilde{Y}\Big|_{\imath(\mathcal{U})}$ is independent of the extension, and depends only on X, Y.

Now, we can uniquely decompose:

$$\nabla_{\widetilde{X}} \widetilde{Y} = \top \nabla_{\widetilde{X}} \widetilde{Y} + \perp \nabla_{\widetilde{X}} \widetilde{Y}$$

where $\top$ is the tangential component and $\perp$ the normal.

Theorem 3.1. *i) Let $D : \mathfrak{X}(\Sigma) \times \mathfrak{X}(\Sigma) \to \mathfrak{X}(\Sigma)$ be defined by:*

$$\imath_*(D_X Y) = \top \nabla_{\widetilde{X}} \widetilde{Y}$$

for all $X, Y \in \mathfrak{X}(\Sigma)$, where $\widetilde{X}, \widetilde{Y}$ are any (local) extensions of X, Y. Then D is the Levi-Civita connection of the induced metric h. (Note that the formula above determines $D_X Y$ uniquely by the infectivity of $\imath_*$).

ii) Let $k : \mathfrak{X}(\Sigma) \times \mathfrak{X}(\Sigma) \to C^\infty(\Sigma;\mathbb{R})$ *be defined by:*

$$k(X,Y) := \imath^* \left[g\left(N, \nabla_{\widetilde{X}} \widetilde{Y}\right)\right],$$

where $\widetilde{X}, \widetilde{Y}$ *are any (local) extensions of* X, Y*. We have that*

$$k(X,Y) = k(Y,X)$$

and

$$k(fX,Y) = fk(X,Y) \qquad \forall \quad f \in C^\infty(\Sigma, \mathbb{R}).$$

Proof. i) We first verify that D is a connection. By the linearity of the orthogonal projection and the push-forward map, we have

$$\begin{aligned}
\imath_*\left(D_{X_1+X_2}Y\right) &= \top \nabla_{\widetilde{X}_1+\widetilde{X}_2}\widetilde{Y} \\
&= \top\left(\nabla_{\widetilde{X}_1}\widetilde{Y} + \nabla_{\widetilde{X}_1}\widetilde{Y}\right) \\
&= \top\nabla_{\widetilde{X}_1}\widetilde{Y} + \top\nabla_{\widetilde{X}_1}\widetilde{Y} \\
&= \imath_*(D_{X_1}Y) + \imath_*(D_{X_2}Y) \\
&= \imath_*(D_{X_1}Y + D_{X_2}Y)
\end{aligned}$$

so by the injectivity of $\imath_*$, we have $D_{X_1+X_2}Y = D_{X_1}Y + D_{X_2}Y$. A similar calculation shows:

$$D_X(Y_1+Y_2) = D_XY_1 + D_XY_2.$$

We have to check the rules for $D_{fX}Y$ and $D_X(fY)$ hold. We note that by Corollary A.3 any $f \in C^\infty(\Sigma;\mathbb{R})$ can locally be written as $f = \imath^*\widetilde{f}$ for some $\widetilde{f} \in C^\infty(\mathcal{M};\mathbb{R})$. We calculate:

$$\begin{aligned}
\imath_*(D_{fX}Y) &= \top\nabla_{\widetilde{f}\widetilde{X}}\widetilde{Y} \\
&= \top\left(\widetilde{f}\nabla_{\widetilde{X}}\widetilde{Y}\right) \\
&= \widetilde{f}\top\nabla_{\widetilde{X}}\widetilde{Y} = \imath^*\left(fD_XY\right)
\end{aligned}$$

similarly

$$\begin{aligned}
\imath_*(D_X[fY]) &= \top\nabla_{\widetilde{X}}\widetilde{f}\widetilde{Y} \\
&= \top\left(\widetilde{f}\nabla_{\widetilde{X}}\widetilde{Y} + \widetilde{X}(\widetilde{f})\widetilde{Y}\right) \\
&= \top\left(\widetilde{f}\nabla_{\widetilde{X}}\widetilde{Y}\right) + \widetilde{X}(\widetilde{f})Y^* \\
&= \imath^*\left(fD_XY + X(f)Y\right)
\end{aligned}$$

Hence D is an affine connection. It remains to show that it is torsion free and metric. To verify that D is torsion free, we calculate

$$\begin{aligned}
\imath_*\left(D_XY - D_YX - [X,Y]\right) &= \top\left(\nabla_{\widetilde{X}}\widetilde{Y} - \nabla_{\widetilde{Y}}\widetilde{X}\right) - [X,Y]^* \\
&= \top\left(\nabla_{\widetilde{X}}\widetilde{Y} - \nabla_{\widetilde{Y}}\widetilde{X}\right) - \left[\widetilde{X},\widetilde{Y}\right] \\
&= \top\left(\nabla_{\widetilde{X}}\widetilde{Y} - \nabla_{\widetilde{Y}}\widetilde{X} - \left[\widetilde{X},\widetilde{Y}\right]\right) = 0.
\end{aligned}$$

Here we have used Lemma A.9. Finally, to check that D respects the induced metric h, we calculate:

$$\begin{aligned} X[h(Y,Z)] &= X[\imath^*(g(\widetilde{Y},\widetilde{Z}))] = \imath^*[\widetilde{X}g(\widetilde{Y},\widetilde{Z})] \\ &= \imath^*\left[g(\nabla_{\widetilde{X}}\widetilde{Y},\widetilde{Z}) + g(\widetilde{Y},\nabla_{\widetilde{X}}\widetilde{Z})\right] \\ &= \imath^*\left[g(\top\nabla_{\widetilde{X}}\widetilde{Y},\widetilde{Z}) + g(\widetilde{Y},\top\nabla_{\widetilde{X}}\widetilde{Z})\right] \\ &= \imath^*\left[g(\imath_*(D_XY),\widetilde{Z}) + g(\widetilde{Y},\imath_*(D_XZ))\right] \\ &= h\left(D_XY,Z\right) + h\left(Y,D_XZ\right). \end{aligned}$$

ii) We know by Lemma A.9 that $[\widetilde{X},\widetilde{Y}]$ is an extension of $[X,Y]$, so we have $[\widetilde{X},\widetilde{Y}]$ is tangent to $\imath(\Sigma)$, hence $g\left(N,[\widetilde{X},\widetilde{Y}]\right) = 0$. Thus:

$$0 = g\left(N,\nabla_{\widetilde{X}}\widetilde{Y} - \nabla_{\widetilde{Y}}\widetilde{X} - [\widetilde{X},\widetilde{Y}]\right) = g\left(N,\nabla_{\widetilde{X}}\widetilde{Y}\right) - g\left(N,\nabla_{\widetilde{Y}}\widetilde{X}\right)$$

so that

$$k(X,Y) = \imath^*\left[g\left(N,\nabla_{\widetilde{X}}\widetilde{Y}\right)\right] = \imath^*\left[g\left(N,\nabla_{\widetilde{Y}}\widetilde{X}\right)\right] = k(Y,X).$$

To establish the linearity, we calculate:

$$\begin{aligned} k(fX,Y) &= \imath^*\left[g\left(N,\nabla_{\widetilde{fX}}\widetilde{Y}\right)\right] \\ &= \imath^*\left[\widetilde{f}g\left(N,\nabla_{\widetilde{X}}\widetilde{Y}\right)\right] \\ &= \left(\imath^*\widetilde{f}\right)\imath^*\left[g\left(N,\nabla_{\widetilde{X}}\widetilde{Y}\right)\right] = fk(X,Y). \end{aligned}$$

Which establishes the result. □

From the second part of this theorem, we deduce that k is a $(0,2)-$tensor field defined on Σ, known as the *second fundamental form*. Notice that since $g(\widetilde{Y},N) = 0$ on $\imath(\Sigma)$, we must have that

$$\begin{aligned} 0 &= \widetilde{X}\left[g(\widetilde{Y},N)\right]\Big|_{\imath(\Sigma)} \\ &= \left[g\left(\nabla_{\widetilde{X}}N,\widetilde{Y}\right) + g\left(N,\nabla_{\widetilde{X}}\widetilde{Y}\right)\right]\Big|_{\imath(\Sigma)} \end{aligned}$$

So that we have:

$$k(X,Y) = -\imath^*\left[g\left(\nabla_{\widetilde{X}}N,\widetilde{Y}\right)\right], \tag{3.9}$$

which is known as *Weingarten's equation*, and gives an alternative approach to finding k.

Example 13. Suppose $\mathcal{M} = \mathbb{R}^4$ with coordinates $(t,x^i)_{i=1,2,3}$, and suppose that g is a Lorentzian metric on $\mathcal{M}$ given by:

$$g = -\phi(t,x)^2dt^2 + h_{ij}(t,x)dx^idx^j,$$

where $\phi > 0$ and h_{ij} is symmetric and positive definite for all (t, x). Consider $\Sigma = \mathbb{R}^3$ with coordinates $(y^i)_{i=1,2,3}$ and consider the map

$$\begin{array}{rccl} \imath : & \Sigma & \hookrightarrow & \mathcal{M} \\ & (y^i) & \mapsto & (0, y^i). \end{array}$$

So that $\imath(\Sigma) = \{t = 0\}$. Now, we note that if $f \in C^\infty(\mathcal{M}; \mathbb{R})$, then $\imath^* f(y) = f(0, y^i)$, so that pulling back a function from $\mathcal{M}$ to Σ simply consists of restricting f to $\{t = 0\}$. By considering the coordinate curves, we can also see that on $\{t = 0\}$ we have:

$$\imath_* \frac{\partial}{\partial y^i} = \frac{\partial}{\partial x^i}.$$

This suggests that the vector fields $\frac{\partial}{\partial x^i}$ are a suitable extension of $\frac{\partial}{\partial y^i}$.

Considering the pull-back of g to Σ, we find:

$$h\left(\frac{\partial}{\partial y^i}, \frac{\partial}{\partial y^j}\right)\bigg|_y = g\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right)\bigg|_{\imath(y)} = h_{ij}(0, y)$$

so that

$$h := \imath^* g = h_{ij}(0, y) dy^i dy^j.$$

To find the second fundamental form, we first note that the future directed unit normal is:

$$N = \frac{1}{\phi}\frac{\partial}{\partial t},$$

which again admits an obvious extension away from $\{t = 0\}$. We calculate the second fundamental form as:

$$\begin{aligned} k\left(\frac{\partial}{\partial y^i}, \frac{\partial}{\partial y^j}\right)\bigg|_y &= g\left(N, \nabla_{\frac{\partial}{\partial x^i}} \frac{\partial}{\partial x^j}\right)\bigg|_{\imath(y)} \\ &= g\left(\frac{1}{\phi}\frac{\partial}{\partial t}, \Gamma^\mu{}_{ij}\frac{\partial}{\partial x^\mu}\right)\bigg|_{\imath(y)} \\ &= -\phi \Gamma^t{}_{ij}\big|_{\imath(y)} \\ &= -\frac{\phi}{2} g^{t\mu}\left(\frac{\partial g_{i\mu}}{\partial x^j} + \frac{\partial g_{j\mu}}{\partial x^i} - \frac{\partial g_{ij}}{\partial x^\mu}\right)\bigg|_{\imath(y)} \\ &= -\frac{1}{2\phi(0, y)}\frac{\partial h_{ij}}{\partial t}(0, y). \end{aligned}$$

Thus the second fundamental form represents the 'first time derivative' of the induced metric. We might expect that the induced metric and the second fundamental form would represent the correct 'Cauchy data' for Einstein's equations, and we shall indeed see that this is the case.

3.3.2 The Gauss and Codazzi-Mainardi equations

The induced metric and second fundamental form carry information about how the surface Σ is 'glued into' the Lorentzian manifold $(\mathcal{M}, g)$. If the manifold satisfies some equations (for example Einstein's equations) then we expect that this is reflected in the information induced by g on Σ by the embedding map $\imath$. We shall see in this section that certain components of the curvature of g at $\imath(\Sigma)$ can be written in terms of h and k. This in turn will imply that when we impose conditions on g, this will be reflected as conditions on h, k.

Theorem 3.2 (Gauss' Equation)**.** *Let $(\mathcal{M}, g)$ be a smooth, time oriented spacetime, and let Σ be a three-dimensional manifold. Suppose that $\imath : \Sigma \hookrightarrow \mathcal{M}$ is an embedding of Σ such that $\imath(\Sigma)$ is spacelike. Suppose $\mathcal{U} \subset \Sigma$ is open. Let $X, Y, Z, W \in \mathfrak{X}(\mathcal{U})$ have extensions $\widetilde{X}, \widetilde{Y}, \widetilde{Z}, \widetilde{W} \in \mathfrak{X}(\mathcal{M})$ away from $\imath(\mathcal{U})$. Then:*

$$\begin{aligned}\imath^* \left[g\left({}^\nabla R(\widetilde{X}, \widetilde{Y})\widetilde{Z}, \widetilde{W}\right)\right] = h\left({}^D R(X, Y)Z, W\right) \\ - k(X, Z)k(Y, W) + k(X, W)k(Y, Z),\end{aligned} \tag{3.10}$$

holds in $\mathcal{U}$, where ${}^\nabla R$, ${}^D R$ are the curvature operators corresponding to ∇, D respectively.

Proof. We will use the splitting of the connection ∇ induced by the embedding, as described in Theorem 3.1.

1. First note that
$$\nabla_{\widetilde{Y}} \widetilde{Z} = \top \nabla_{\widetilde{Y}} \widetilde{Z} - g(\nabla_{\widetilde{Y}} \widetilde{Z}, N)N. \tag{3.11}$$
Replacing $\widetilde{Y}$ with $\left[\widetilde{X}, \widetilde{Y}\right]$, and taking the inner product with $\widetilde{W}$, we have that on $\imath(\mathcal{U})$:
$$g\left(\nabla_{[\widetilde{X}, \widetilde{Y}]} \widetilde{Z}, \widetilde{W}\right) = g\left(\top \nabla_{[\widetilde{X}, \widetilde{Y}]} \widetilde{Z}, \widetilde{W}\right) = g\left(\imath_* D_{[X,Y]} Z, \imath_* W\right)$$
so that
$$\imath^* \left[g\left(\nabla_{[\widetilde{X}, \widetilde{Y}]} \widetilde{Z}, \widetilde{W}\right)\right] = h\left(D_{[X,Y]} Z, W\right) \tag{3.12}$$

2. Differentiating (3.11) in the $\widetilde{X}$ direction and acting with $\top$, we have:
$$\top\left(\nabla_{\widetilde{X}} \nabla_{\widetilde{Y}} \widetilde{Z}\right) = \top\left(\nabla_{\widetilde{X}} \top \nabla_{\widetilde{Y}} \widetilde{Z}\right) - g(\nabla_{\widetilde{Y}} \widetilde{Z}, N)\nabla_{\widetilde{X}} N + \widetilde{X}\left[g(\nabla_{\widetilde{Y}} \widetilde{Z}, N)\right] N$$
We note that on $\imath(\mathcal{U})$, the first term is equal to $\imath_* D_X D_Y Z$, by Theorem 3.1 *i*), and the last term is in the normal direction. Now, let us take the inner product with $\widetilde{W}$ and pull-back by $\imath$ to obtain:
$$\begin{aligned}\imath^* \left[g\left(\nabla_{\widetilde{X}} \nabla_{\widetilde{Y}} \widetilde{Z}, \widetilde{W}\right)\right] &= \imath^* \left[g\left(\top\left(\nabla_{\widetilde{X}} \nabla_{\widetilde{Y}} \widetilde{Z}\right), \widetilde{W}\right)\right] \\ &= \imath^* \left[g\left(\top\left(\nabla_{\widetilde{X}} \top \nabla_{\widetilde{Y}} \widetilde{Z}\right), \widetilde{W}\right)\right] - \imath^* \left[g(\nabla_{\widetilde{Y}} \widetilde{Z}, N) g\left(N, \nabla_{\widetilde{X}}, \widetilde{W}\right)\right] \\ &= h(D_X D_Y Z, W) + k(Y, Z)k(X, W)\end{aligned} \tag{3.13}$$
Here, we have used the definition of k from Theorem 3.1 *i*), together with Weingarten's equation (3.9) to deal with the second term on the right hand side.

3. Now, we simply use the definition of ${}^{\nabla}R$:

$$ {}^{\nabla}R(\widetilde{X},\widetilde{Y})\widetilde{Z} = \nabla_{\widetilde{X}}\nabla_{\widetilde{Y}}\widetilde{Z} - \nabla_{\widetilde{Y}}\nabla_{\widetilde{X}}\widetilde{Z} - \nabla_{[\widetilde{X},\widetilde{Y}]}\widetilde{Z}. $$

Taking the inner product of this equation with $\widetilde{W}$, and pulling back by $\imath$, we deduce:

$$ \begin{aligned} \imath^*\left[g\left({}^{\nabla}R(\widetilde{X},\widetilde{Y})\widetilde{Z},\widetilde{W}\right)\right] &= \imath^*\left[g\left(\nabla_{\widetilde{X}}\nabla_{\widetilde{Y}}\widetilde{Z},\widetilde{W}\right)\right] - \imath^*\left[g\left(\nabla_{\widetilde{Y}}\nabla_{\widetilde{X}}\widetilde{Z},\widetilde{W}\right)\right] \\ &\quad - \imath^*\left[g\left(\nabla_{[\widetilde{X},\widetilde{Y}]}\widetilde{Z},\widetilde{W}\right)\right] \\ &= h(D_X D_Y Z, W) + k(Y,Z)k(X,W) \\ &\quad - h(D_Y D_Y X, W) - k(X,Z)k(Y,W) \\ &\quad - h\left(D_{[X,Y]}Z, W\right) \\ &= h\left({}^{D}R(X,Y)Z, W\right) \\ &\quad - k(X,Z)k(Y,W) + k(X,W)k(Y,Z), \end{aligned} $$

where we have used (3.12), (3.13) to pass from the first equality to the second equality. This is the result we require. □

Notice that the right hand side of Gauss' equation involves only geometric objects defined on the surface Σ. The left hand side is a quantity defined on the full spacetime. Note also that while all of the components of the Riemann tensor of h can appear on the right hand side, on the left hand side we can only realise purely tangential components of the Riemann tensor of g.

Gauss' equation tells us that the curvature of $(\mathcal{M}, g)$ in tangential directions to $\imath(\Sigma)$ is reflected both in the intrinsic curvature of Σ, thought of as a Riemannian manifold with metric h, as well as in the second fundamental form, which is sometimes referred to as the extrinsic curvature. This difference between intrinsic and extrinsic curvature is an important distinction even in the study of surfaces in $\mathbb{R}^3$. For example, a cylinder in $\mathbb{R}^3$ has no intrinsic curvature (the induced metric is flat) but it does have extrinsic curvature.

We shall also require another result relating the curvature of g to the intrinsic data h, k on Σ. This result involves components of the Riemann tensor of g which are not entirely tangential, but involve contraction with a normal direction.

Theorem 3.3 (Codazzi-Mainardi equation)**.** *Let $(\mathcal{M}, g)$ be a smooth, time oriented spacetime, and let Σ be a three-dimensional manifold. Suppose that $\imath : \Sigma \hookrightarrow \mathcal{M}$ is an embedding of Σ such that $\imath(\Sigma)$ is spacelike and take $\mathcal{U} \subset \Sigma$ an open subset. Let $X, Y, Z \in \mathfrak{X}(\mathcal{U})$ have extensions $\widetilde{X}, \widetilde{Y}, \widetilde{Z} \in \mathfrak{X}(\mathcal{M})$ away from $\imath(\mathcal{U})$ and suppose that $N \in \mathfrak{X}(\mathcal{M})$ agrees with the future directed unit normal on $\imath(\mathcal{U})$. Then:*

$$ \imath^*\left[g\left({}^{\nabla}R(\widetilde{X},\widetilde{Y})\widetilde{Z}, N\right)\right] = [D_X k](Y,Z) - [D_Y k](X,Z), \tag{3.14} $$

holds in $\mathcal{U}$

Proof. 1. Recall (3.11) from the proof of Gauss' equation

$$ \nabla_{\widetilde{Y}}\widetilde{Z} = \top\nabla_{\widetilde{Y}}\widetilde{Z} - g(\nabla_{\widetilde{Y}}\widetilde{Z}, N)N. \tag{3.15} $$

Replacing $\widetilde{Y}$ with $\left[\widetilde{X}, \widetilde{Y}\right]$, and taking the inner product with N, we have that on $\imath(\mathcal{U})$:

$$g\left(\nabla_{[\widetilde{X},\widetilde{Y}]}\widetilde{Z}, N\right) = g(\nabla_{[\widetilde{X},\widetilde{Y}]}\widetilde{Z}, N)$$

so that

$$\imath^*\left[g\left(\nabla_{[\widetilde{X},\widetilde{Y}]}\widetilde{Z}, N\right)\right] = k\left([X,Y], Z\right) = k(D_X Y - D_Y X, Z) \tag{3.16}$$

where we use the definition of k from Theorem 3.1 and the fact that D is torsion-free.

2. Differentiating (3.15) in the $\widetilde{X}$ direction and forming the inner product with N, we have:

$$\begin{aligned} g\left(\nabla_{\widetilde{X}}\nabla_{\widetilde{Y}}\widetilde{Z}, N\right) = g\left(\nabla_{\widetilde{X}}{}^{\top}\nabla_{\widetilde{Y}}\widetilde{Z}, N\right) - g\left(\nabla_{\widetilde{Y}}\widetilde{Z}, N\right) g\left(\nabla_{\widetilde{X}}N, N\right) \\ + \widetilde{X}\left[g\left(\nabla_{\widetilde{Y}}\widetilde{Z}, N\right)\right] \end{aligned} \tag{3.17}$$

Now, note that since $g(N,N) = -1$ on $\imath(\Sigma)$, and $\tilde{X}$ is tangent to $\imath(\Sigma)$, we have

$$0 = \widetilde{X}\left[g(N,N)\right]\Big|_{\imath(\Sigma)} = 2g\left(\nabla_{\widetilde{X}}N, N\right)\big|_{\imath(\Sigma)},$$

so that the second term on the right of (3.17) vanishes on $\imath(\Sigma)$. Pulling (3.17) back by $\imath$, we thus have:

$$\imath^*\left[g\left(\nabla_{\widetilde{X}}\nabla_{\widetilde{Y}}\widetilde{Z}, N\right)\right] = k\left(D_Y Z, X\right) + X\left[k(Y,Z)\right] \tag{3.18}$$

3. Now, we use the definition of ${}^{\nabla}R$:

$${}^{\nabla}R(\widetilde{X},\widetilde{Y})\widetilde{Z} = \nabla_{\widetilde{X}}\nabla_{\widetilde{Y}}\widetilde{Z} - \nabla_{\widetilde{Y}}\nabla_{\widetilde{X}}\widetilde{Z} - \nabla_{[\widetilde{X},\widetilde{Y}]}\widetilde{Z}.$$

Taking the inner product of this equation with N, and pulling back by $\imath$, we deduce:

$$\begin{aligned} \imath^*\left[g\left({}^{\nabla}R(\widetilde{X},\widetilde{Y})\widetilde{Z}, N\right)\right] &= \imath^*\left[g\left(\nabla_{\widetilde{X}}\nabla_{\widetilde{Y}}\widetilde{Z}, N\right)\right] - \imath^*\left[g\left(\nabla_{\widetilde{Y}}\nabla_{\widetilde{X}}\widetilde{Z}, N\right)\right] \\ &\quad - \imath^*\left[g\left(\nabla_{[\widetilde{X},\widetilde{Y}]}\widetilde{Z}, N\right)\right] \\ &= k\left(D_Y Z, X\right) + X\left[k(Y,Z)\right] \\ &\quad - k\left(D_X Z, Y\right) - Y\left[k(Y,Z)\right] \\ &\quad - k(D_X Y - D_Y X, Z) \\ &= X\left[k(Y,Z)\right] - k\left(D_X Z, Y\right) - k(Z, D_X Y) \\ &\quad - \left(Y\left[k(Y,Z)\right] - k\left(D_Y Z, X\right) - k(Z, D_Y X)\right) \\ &= \left[D_X k\right](Y,Z) - \left[D_Y k\right](X,Z). \end{aligned}$$

Here we have used (3.16), (3.18) to pass from the first inequality to the second, and the definition of $\nabla_X k$, $\nabla_Y k$ for the final equality. This is the expression we require. □

The calculations in this section are somewhat involved, but the basic idea is to use and reuse the splitting of the connection into tangential and normal parts that we discussed in Theorem 3.1. The main things to take away are the equations (3.10), (3.14) which allow us to relate certain components of the curvature of g to the objects induced on Σ by $\imath$, namely h and k.

3.3.3 The Einstein constraint equations

Now, since we have expressed certain components of the Riemann tensor of g in terms of the quantities h, k, it is natural that imposing conditions on the Riemann tensor of g will impose conditions on h and k. In particular, if we assume that the metric g satisfies Einstein's equations, we shall see that certain relations must hold between h and k. These are the *Einstein constraint equations.*

Let us suppose $\mathcal{U} \subset \Sigma$ is open and that $\{e_i\}_{i=1,2,3}$, where $e_i \in \mathfrak{X}(\mathcal{U})$, form a local basis which is orthonormal with respect to h. We can assume that there exist $\widetilde{e}_i \in \mathfrak{X}(\mathcal{M})$ which extend e_i away from $\imath(\mathcal{U})$. Setting $\widetilde{e}_0 = N$, we have that $\{\widetilde{e}_\mu\}_{\mu=0,\ldots,3}$ is a basis on some neighbourhood of $\imath(\mathcal{U}) \subset \mathcal{M}$, which is orthonormal at each point of $\imath(\mathcal{U})$.

Let us now take traces of the Gauss equation (3.10) to relate the Ricci curvature of g to quantities defined on Σ. We have:

$$
\begin{aligned}
\imath^* \left[g\left({}^{\nabla}R(\widetilde{e}_i, \widetilde{Y})\widetilde{Z}, \widetilde{e}_j\right)\delta^{ij}\right] &= \imath^* \left[g\left({}^{\nabla}R(\widetilde{e}_\mu, \widetilde{Y})\widetilde{Z}, \widetilde{e}_\nu\right)\eta^{\mu\nu}\right] + \imath^* \left[g\left({}^{\nabla}R(\widetilde{e}_0, \widetilde{Y})\widetilde{Z}, \widetilde{e}_0\right)\right] \\
&= \imath^* \left[Ric_g(\widetilde{Y}, \widetilde{Z})\right] + \imath^* \left[g\left({}^{\nabla}R(\widetilde{e}_0, \widetilde{Y})\widetilde{Z}, \widetilde{e}_0\right)\right].
\end{aligned}
$$

On the other hand, we have

$$
\begin{aligned}
\imath^* \left[g\left({}^{\nabla}R(\widetilde{e}_i, \widetilde{Y})\widetilde{Z}, \widetilde{e}_j\right)\delta^{ij}\right] &= \delta^{ij} h\left({}^{D}R(e_i, Y)Z, e_j\right) \\
&\qquad - \delta^{ij} k(e_i, Z) k(e_j, Y) + \delta^{ij} k(e_i, e_j) k(Y, Z) \\
&= Ric_h(Y, Z) - \delta^{ij} k(e_i, Z) k(e_j, Y) + (\mathrm{Tr}_h k)\, k(Y, Z)
\end{aligned}
$$

So that:

$$
\begin{aligned}
&\imath^* \left[Ric_g(\widetilde{Y}, \widetilde{Z})\right] + \imath^* \left[g\left({}^{\nabla}R(\widetilde{e}_0, \widetilde{Y})\widetilde{Z}, \widetilde{e}_0\right)\right] \\
&\qquad = Ric_h(Y, Z) - \delta^{ij} k(e_i, Z) k(e_j, Y) + (\mathrm{Tr}_h k)\, k(Y, Z)
\end{aligned}
$$

Einstein's equations don't impose a condition on the left hand side of this equation, so we trace again over the Y, Z slots. We have:

$$
\begin{aligned}
&\imath^* \left[Ric_g(\widetilde{e}_k, \widetilde{e}_l)\right]\delta^{kl} + \imath^* \left[g\left({}^{\nabla}R(\widetilde{e}_0, \widetilde{e}_k)\widetilde{e}_l, \widetilde{e}_0\right)\delta^{kl}\right] \\
&\qquad = \imath^* \left[Ric_g(\widetilde{e}_k, \widetilde{e}_l)\right]\delta^{kl} + \imath^* \left[g\left({}^{\nabla}R(\widetilde{e}_0, \widetilde{e}_\sigma)\widetilde{e}_\tau, \widetilde{e}_0\right)\eta^{\sigma\tau}\right] + \imath^* \left[g\left({}^{\nabla}R(\widetilde{e}_0, \widetilde{e}_0)\widetilde{e}_0, \widetilde{e}_0\right)\right] \\
&\qquad = \imath^* \left[Ric_g(\widetilde{e}_k, \widetilde{e}_l)\right]\delta^{kl} + \imath^* \left[Ric_g(\widetilde{e}_0, \widetilde{e}_0)\right] \\
&\qquad = \imath^* \left[R_g + 2Ric_g(\widetilde{e}_0, \widetilde{e}_0)\right],
\end{aligned}
$$

where we have made use of the symmetries of the Riemann tensor in several places. We also have:

$$Ric_h(e_k, e_k)\delta^{kl} - \delta^{ij}\delta^{kl}k(e_i, e_k)k(e_j, e_l) + (\mathrm{Tr}_h k)\, k(e_k, e_l)\delta^{kl}$$
$$= R_h - |k|_h^2 + (\mathrm{Tr}_h k)^2$$

Let us suppose that g satisfies Einstein's equations, with some energy-momentum tensor T. Notice that on $\imath(\Sigma)$:

$$R_g + 2Ric_g(\widetilde{e}_0, \widetilde{e}_0) = 2\left(Ric_g - \frac{1}{2}R_g g\right)(\tilde{e}_0, \tilde{e}_0) = 2T(\tilde{e}_0, \tilde{e}_0) + 2\Lambda.$$

Now, $T(\tilde{e}_0, \tilde{e}_0)$ is the energy density of the matter fields, as measured by an observer whose instantaneous velocity is given by $\widetilde{e}_0$. We introduce $\rho \in C^\infty(\Sigma; \mathbb{R})$, the local energy density of matter fields on Σ, defined by:

$$\rho := \imath^*\left[T(\tilde{e}_0, \tilde{e}_0)\right].$$

Then, putting everything together, we have the first Einstein constraint equation:

$$R_h - |k|_h^2 + (\mathrm{Tr}_h k)^2 = 2\rho + 2\Lambda \tag{3.19}$$

In particular, in the vacuum case with vanishing cosmological constant, we have:

$$R_h - |k|_h^2 + (\mathrm{Tr}_h k)^2 = 0.$$

We can perform a similar procedure with the Codazzi-Mainardi equation (3.14). We have:

$$\imath^*\left[g\left({}^\nabla R(\widetilde{X}, \widetilde{e}_i)\widetilde{e}_j, \widetilde{e}_0\right)\right]\delta^{ij} = \imath^*\left[g\left({}^\nabla R(\widetilde{X}, \widetilde{e}_\mu)\widetilde{e}_\nu, \widetilde{e}_0\right)\eta^{\mu\nu}\right] + \imath^*\left[g\left({}^\nabla R(\widetilde{X}, \widetilde{e}_0)\widetilde{e}_0, \widetilde{e}_0\right)\right]$$
$$= \imath^*\left[Ric_g\left(\widetilde{X}, \widetilde{e}_0\right)\right]$$

We also have:

$$\left[D_X k\right](e_i, e_j)\delta^{ij} - \left[D_{e_i} k\right](X, e_j)\delta^{ij} = X\left[\mathrm{Tr}_h K\right] - \mathrm{div}_h k\,(X),$$

where we are using that D is the Levi-Civita connection of h and the fact that e_i is an orthonormal basis for h. Since $\widetilde{e}_0$ is normal to $\imath(\Sigma)$, we have that on $\imath(\Sigma)$:

$$Ric_g\left(\widetilde{X}, \widetilde{e}_0\right) = \left(Ric_g - \frac{1}{2}R_g g\right)(\widetilde{X}, \widetilde{e}_0) = T(\widetilde{X}, \widetilde{e}_0).$$

Now $-T(\cdot, \widetilde{e}_0)$, which we think of as a one-form, is the momentum flux density of the matter fields, as measured by an observer whose instantaneous velocity is given by $\widetilde{e}_0$. We introduce $J \in \mathfrak{X}^*(\Sigma)$, the local momentum flux density of matter fields on Σ, defined by:

$$J := -\imath^*\left[T(\cdot, \widetilde{e}_0)\right].$$

Putting this together with the Codazzi-Mainardi equation, we deduce the second Einstein constraint equation:

$$\mathrm{div}_h k - d\,(\mathrm{Tr}_h K) = J.$$

In particular, in the vacuum case with vanishing cosmological constant, we have:

$$\mathrm{div}_h k - d\,(\mathrm{Tr}_h K) = 0.$$

Exercise 3.7. Recall the Schwarzschild metric in Painlevé-Gullstrand coordinates (Examples 7, 9, 12) is a Lorentzian metric on $\mathcal{M} = \mathbb{R} \times (0, \infty) \times S^2$ given by:

$$g = -\left(1 - \frac{2m}{r}\right) dt^2 + 2\sqrt{\frac{2m}{r}} dt dr + dr^2 + r^2 g_{S^2},$$

And let us choose the local basis of vector fields $\{e_\mu\}$, as in Examples 7, 9. Let $\Sigma = (0, \infty) \times S^2 \simeq \mathbb{R}^3 \setminus \{0\}$. Consider the map

$$\begin{array}{rccc} \imath : & \Sigma & \hookrightarrow & \mathcal{M} \\ & (\rho, \boldsymbol{X}) & \mapsto & (0, \rho, \boldsymbol{X}). \end{array}$$

In other words, the suface $\imath(\Sigma)$ is the surface $\{t = 0\}$.

a) Show that the future directed unit normal of $\imath(\Sigma)$ is given by:

$$N = e_0 = \frac{\partial}{\partial t} - \sqrt{\frac{2m}{r}} \frac{\partial}{\partial r}.$$

b) Show that the induced metric h is the canonical flat metric on $\mathbb{R}^3 \setminus \{0\}$:

$$h = d\rho^2 + \rho^2 g_{S^2}.$$

c) Show that:

$$k = -\frac{1}{2\rho}\sqrt{\frac{2m}{\rho}} d\rho^2 + \rho\sqrt{\frac{2m}{\rho}} g_{S^2}.$$

[Hint: consider $k(b_i, b_j)$ for a suitable basis of vector fields on Σ such that $\imath^* b_i = e_i$, and use (2.13)]

d*) Under a change of coordinates $\boldsymbol{x} = \rho \boldsymbol{X}$ from polar to Cartesian coordinates, you are given that h and k become:

$$h = \delta_{ij} dx^i dx^j,$$
$$k = \sqrt{\frac{2m}{|\boldsymbol{x}|^3}} \left(\delta_{ij} - \frac{3}{2} \frac{x_i x_j}{|\boldsymbol{x}|^2}\right) dx^i dx^j.$$

Show that:

$$|k|_h^2 - (\mathrm{Tr}_h k)^2 = 0,$$

and

$$\mathrm{div}_h k - d\,(\mathrm{Tr}_h k) = 0.$$

[Hint: Note that since h is the canonical metric on $\mathbb{R}^3$ in Cartesian coordinates, $(\mathrm{div}_h k)_j = \partial_i k_{ij}$ and $(df)_i = \partial_i f$.]

3.4 The Cauchy problem

For the purposes of studying the vacuum Einstein equations, we can summarise the results of the previous section in the following theorem:

Theorem 3.4. *Let $(\mathcal{M}, g)$ be a smooth, time oriented spacetime satisfying the vacuum Einstein equations with vanishing cosmological constant:*

$$Ric_g = 0.$$

Suppose that Σ is a smooth three dimensional manifold and that $\imath : \Sigma \hookrightarrow \mathcal{M}$ is a smooth embedding whose image is everywhere spacelike. Then the metric h and second fundamental form k induced on Σ by $\imath$ satisfy the Einstein constraint equations:

$$0 = R_h - |k|_h^2 + (\mathrm{Tr}_h k)^2 , \tag{3.20}$$

$$0 = \mathrm{div}_h k - d\, (\mathrm{Tr}_h k) . \tag{3.21}$$

We want to consider the Cauchy problem for Einstein's equations. Loosely, we wish to specify data on some initial hypersurface, and then construct a solution which represents the evolution of that data into the future. Recall that in the case of a field ψ satisfying the wave equation, the correct Cauchy data on a spacelike hypersurface, Σ, was $\psi|_\Sigma$ and $N_\Sigma \psi|_\Sigma$. In the case of Einstein's equations, a natural candidate to take the place of $\psi|_\Sigma$ is h, the induced metric. By considering Example 13, we can see that a natural candidate to take the place of $N_\Sigma \psi|_\Sigma$ is k, the second fundamental form. Theorem 3.4 gives some necessary conditions on h and k such that they represent initial data for Einstein's equations. We shall see that these conditions are in fact sufficient.

Definition 18. An *admissible triple* (Σ, h, k) consists of a smooth $3-$dimensional manifold Σ, equipped with a Riemannian metric h and a symmetric $(0, 2)-$tensor k satisfying the Einstein constraint equations (3.20), (3.21).

Examples of admissible triples include[3] $(\mathbb{R}^3, \delta, 0)$, the data induced on the surface $\{t = 0\}$ in the Minkowski spacetime, as well as the example constructed in Exercise 3.7.

You should think of an admissible triple as giving the 'initial conditions' for Einstein's equations. In contrast to the case of the wave equation, there is a subtlety in defining what we mean by a solution with this initial data. This comes about because we don't know a priori the spacetime manifold $\mathcal{M}$ on which we shall solve Einstein's equations. The correct notion of solution is given by:

Definition 19. Suppose (Σ, h, k) is an admissible triple. A *development* of (Σ, h, k) is a Lorentzian manifold $(\mathcal{M}, g)$, together with an embedding map $\imath : \Sigma \hookrightarrow \mathcal{M}$ such that

i) g satisfies the vacuum Einstein equations in $\mathcal{M}$:

$$Ric_g = 0$$

[3] Here $\delta = \delta_{ij} dx^i dx^j$ is the flat metric on $\mathbb{R}^3$

ii) h is the metric induced by g on Σ under the embedding map $\imath$.

iii) k is the second fundamental form of the embedding $\imath$.

iv) $\mathcal{M}$ is the future Cauchy development of $\imath(\Sigma)$: i.e. $D^+(\Sigma) = \mathcal{M}$.

By 'solving' Einstein's equations with a certain admissible triple as initial data, we mean finding a development of the triple. Before we are able to state the well posedness of Einstein's equations, we need one more ingredient. In order to say that a PDE problem is well posed, we not only require that a solution exists for given initial data, but that moreover it is unique. In the case of Einstein's equations, this is somewhat subtle, because the 'solution' we construct is a geometrical object. We know that the same geometric object can be described in several different ways.

To motivate our next definition, consider $\mathcal{M} = (0, \infty) \times (-\pi, \pi)$ with coordinates (r, θ), endowed with the metric

$$g = dr^2 + r^2 d\theta^2.$$

On the other hand, consider $\mathcal{M}' = \mathbb{R}^2 \setminus \{x \leq 0, y = 0\}$ where (x, y) are the usual coordinates on $\mathbb{R}^2$. We endow $\mathcal{M}'$ with the flat Riemannian metric:

$$g' = dx^2 + dy^2.$$

There is a map between these two manifolds, given by:

$$\begin{array}{rccc} \jmath : & \mathcal{M} & \to & \mathcal{M}' \\ & (r, \theta) & \mapsto & (r \sin\theta, r\cos\theta) \end{array}$$

The map $\jmath$ is smooth, bijective and has smooth inverse, hence $\jmath$ is a diffeomorphism. Moreover, we can verify that

$$\jmath^* g' = g.$$

We say that $\jmath$ is an isometry. Although $(\mathcal{M}, g)$ and $(\mathcal{M}', g')$ are *different* manifolds, we nevertheless think of them as describing the *same* underlying geometry, but in different coordinates.

The solutions that we construct to Einstein's equations will be unique up to transformations of this kind, and extensions.

Definition 20. An isometric embedding from $(\mathcal{M}, g)$ to $(\mathcal{M}', g')$ is an embedding $\jmath : \mathcal{M} \hookrightarrow \mathcal{M}'$ such that $\jmath^* g' = g$.

We are now ready to state the main result of this course:

Theorem 3.5 (Choquet-Bruhat–Geroch)**.** *Given an admissible triple* (Σ, h, k)*, there exists a unique development* $(\mathcal{M}, g, \imath)$ *which is maximal in the sense that if* $(\widetilde{\mathcal{M}}, \widetilde{g}, \widetilde{\imath})$ *is any other development of* (Σ, h, k)*, then there exists an isometric embedding* $\jmath : \widetilde{\mathcal{M}} \hookrightarrow \mathcal{M}$ *such that:*

$$\jmath \circ \tilde{\imath} = \imath.$$

The full proof of this theorem is beyond the scope of the course, but is contained in the book of Choquet-Bruhat[4]. We shall offer a sketch of the proof, which follows a similar structure to our discussion of the linearised problem around Minkowski space.

Sketch proof:

1. We assume there exist global coordinates[5] (x^i) on Σ and extend them to coordinates (t, x^i) on $\mathcal{M} := (-\epsilon, \epsilon) \times \Sigma$. Recall from Theorem 2.9 that Einstein's equations in $\mathcal{M}$ are equivalent to the system of quasilinear wave equations:
$$-\frac{1}{2}g^{\mu\alpha}\frac{\partial^2 g_{\sigma\nu}}{\partial x^\mu \partial x^\alpha} + P_{\sigma\nu}(g, \partial g) = 0 \tag{3.22}$$
provided that the wave coordinate condition holds:
$$\Gamma^\alpha{}_\mu{}^\mu = 0. \tag{3.23}$$

2. By some standard results in the study of nonlinear wave equations, there exists a unique solution to (3.22) with the initial conditions:
$$g|_{t=0} = -dt^2 + h \tag{3.24}$$
$$\partial_t g|_{t=0} = -2k \tag{3.25}$$
provided $\epsilon > 0$ is sufficiently small. Notice that $g|_{t=0}$ is Lorentzian by construction.

3. We define $F^\alpha := \Gamma^\alpha{}_\mu{}^\mu$, and then show that if (3.22) is satisfied, then F^α satisfies a (linear) system of wave equations:
$$\Box_g F^\alpha + (A \cdot F)^\alpha = 0.$$
By similar methods to those used in the proof of Proposition 1 we can show that if $F^\alpha|_{t=0} = \partial_t F^\alpha|_{t=0} = 0$, then $F^\alpha \equiv 0$.

4. We next demonstrate that the condition $F^\alpha|_{t=0} = \partial_t F^\alpha|_{t=0} = 0$, then $F^\alpha \equiv 0$ is equivalent to the constraint equations holding on h, k. By restricting to the future domain of dependence of $\{0\} \times \Sigma$ we have constructed a development of (Σ, h, k).

5. To establish local uniqueness we show that given *any* development of (Σ, h, k) it is possible to construct wave coordinates such that (3.24), (3.25) hold. By the uniqueness of solutions of (3.22), we deduce that given any two developments there is a neighbourhood of the initial surface in each which can mapped onto one another by an isometry.

6. The final stage is to establish the existence of a single *maximal* development. Historically, this was the final part of the result to be established. The issue here is that while we know that in a neighbourhood of Σ two developments are isometric, constructing a larger development in which both embed isometrically is difficult.

[4]"General Relativity and the Einstein Equations", Yvonne Choquet-Bruhat, Oxford 2009. See Chapter VI, §7, 8, 9

[5]This is not a significant restriction by the finite speed of propagation property for hyperbolic PDE.

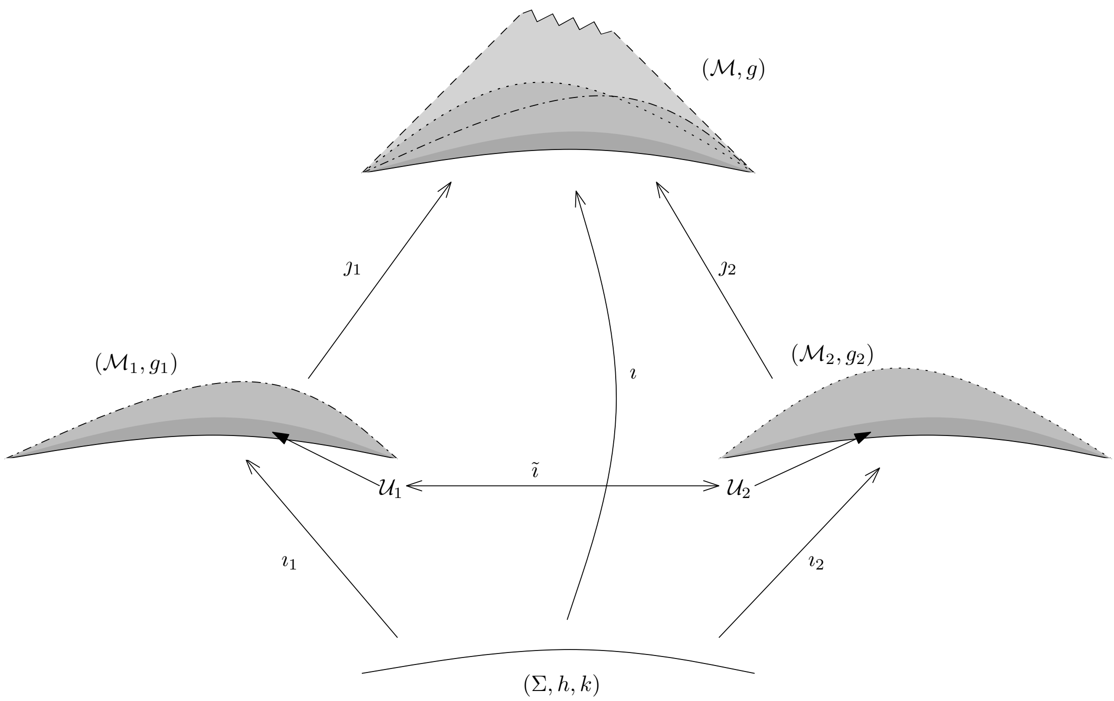

Figure 3.1 Two developments $(\mathcal{M}_i, g_i, \imath_i)$ of an admissible triple, and the maximal development $(\mathcal{M}, g, \imath)$.

The situation is shown in Figure 3.1. Here we have two developments $(\mathcal{M}_1, g_1, \imath_1)$, $(\mathcal{M}_2, g_2, \imath_2)$ of an admissible triple (Σ, h, k). By the previous part we can deduce that there is a neighbourhood $\mathcal{U}_1 \subset \mathcal{M}_1$ of $\imath_1(\Sigma)$, and a neighbourhood $\mathcal{U}_2 \subset \mathcal{M}_2$ of $\imath_2(\Sigma)$ which are mapped isometrically onto one another by $\tilde{\imath}$. The uniqueness theorem states that there exists a larger development, $(\mathcal{M}, g, \imath)$ such that $(\mathcal{M}_i, g_i)$ is isometrically embedded into $(\mathcal{M}, g)$ by the maps $\jmath_i$. Moreover, the map $\jmath_i \circ \imath_i$ which embeds Σ into $\mathcal{M}$ should be the same as $\imath$.

In the original paper of Geroch and Choquet-Bruhat[6] the construction of the maximal development was accomplished by an appeal to Zorn's Lemma (and hence the full Axiom of Choice). Recently the proof has been 'deZornified' by Sbierski[7]. □

While the theorem establishes the existence of a maximal development, it doesn't tell us anything about what the obstacles to extending the solution beyond that development are. For example, the maximal global development of $(\mathbb{R}^3, \delta, 0)$, is the whole of Minkowski space to the future of $\{t = 0\}$: a future complete manifold (i.e., any future directed timelike curve can be extended indefinitely). By contrast, the maximal development of the data constructed in Exercise (3.7) is the region of the Schwarzschild space-time in

[6]"Global aspects of the Cauchy problem in general relativity", Yvonne Choquet-Bruhat, Robert Geroch, Comm. Math. Phys. 14 (1969) p329

[7]"On the Existence of a Maximal Cauchy Development for the Einstein Equations - a Dezornification", Jan Sbierski, arXiv:1309.7591

Painlevé-Gullstrand coordinates to the future of $\{t = 0\}$. This is not future complete, as future directed timelike curves can meet the curvature singularity at $r = 0$.

Much of the current research activity in Mathematical Relativity concerns the global properties of solutions which start from initial data 'close' to the data of an explicitly known solution. For example, one of the crowning achievements of recent work is the following result:

Theorem 3.6 (Christodoulou–Klainerman, '92)**.** *Suppose that we start with an admissible triple* $(\mathbb{R}^3, h, k)$*, with*

$$||h - \delta||_a + ||k||_b < \epsilon,$$

for $\epsilon > 0$ *sufficiently small, where* $||\cdot||_a, ||\cdot||_b$ *are suitable norms*[8]*. Then the maximal development is future complete, and asymptotic to the Minkowski spacetime.*

This result is important because it asserts that for a sufficiently 'small' gravitational field, i.e. a field which is initially sufficiently close to flat space, no singularities form in the evolution. In particular, no black holes are formed.

By contrast, an analogous result is not known for the case of the Schwarzschild black hole discussed above. It is believed that data sufficiently close to the Schwarzschild data constructed in Exercise 3.7 will evolve to give a spacetime containing a region similar to the exterior region, $r > 2m$, of Schwarzschild, however this is at present an unproven conjecture.

[8] In fact, weighted Sobolev norms.

Appendix A

Some background results

A.1 Linear algebra

A.1.1 Vectors and co-vectors

Suppose we have a vector $\boldsymbol{v} \in V$, belonging to an $n-$dimensional real[1] vector space and let $B := \{\boldsymbol{e}_i\}_{i=1,\ldots n}$ be a basis for V. We can write

$$\boldsymbol{v} = \sum_{i=1}^{n} v^i \boldsymbol{e}_i,$$

where $v^i \in \mathbb{R}$ are the uniquely determined components with respect to the basis B. The dual space V^* is the $n-$dimensional real vector space of linear maps $\boldsymbol{\omega} : V \to \mathbb{R}$. Such maps are sometimes called one-forms or covectors. We can define the dual basis $B^* := \{\boldsymbol{e}^i\}_{i=1,\ldots n}$ uniquely by the requirement

$$\boldsymbol{e}^i(\boldsymbol{e}_j) = \delta^i{}_j, \qquad i, j = 1, \ldots, n,$$

where

$$\delta^i{}_j = \begin{cases} 1, & i = j, \\ 0, & i \neq j, \end{cases}$$

is the *Kronecker delta.* For any $\boldsymbol{\omega} \in V^*$ we can write

$$\boldsymbol{\omega} = \sum_{i=1}^{n} \omega_i \boldsymbol{e}^i,$$

[1] for simplicity. Similar constructions exist over other fields.

so that

$$\begin{aligned}\boldsymbol{\omega}(\boldsymbol{v}) &= \sum_{i=1}^{n} \omega_i \boldsymbol{e}^i \left(\sum_{j=1}^{n} v^j \boldsymbol{e}_j \right), \\ &= \sum_{i,j=1}^{n} \omega_i v^j \boldsymbol{e}^i(\boldsymbol{e}_j) = \sum_{i,j=1}^{n} \omega_i v^j \delta^i{}_j, \\ &= \sum_{i=1}^{n} \omega_i v^i.\end{aligned}$$

A.1.2 Tensors

The tensor product

Suppose V, W are real vector spaces of dimensions n, m respectively. From them, we can form a new vector space: the tensor product, denoted $V \otimes W$. The space $V \otimes W$ consists of formal sums of the form

$$\boldsymbol{v}_1 \otimes \boldsymbol{w}_1 + \ldots + \boldsymbol{v}_k \otimes \boldsymbol{w}_k,$$

with $\boldsymbol{v}_p \in V$, $\boldsymbol{w}_p \in W$ for $p = 1, \ldots, k$. The tensor product $\otimes$ is bilinear, obeying:

$$\begin{aligned}\boldsymbol{v} \otimes (\boldsymbol{w}_1 + \lambda \boldsymbol{w}_2) &= \boldsymbol{v} \otimes \boldsymbol{w}_1 + \lambda (\boldsymbol{v} \otimes \boldsymbol{w}_2), \\ (\boldsymbol{v}_1 + \lambda \boldsymbol{v}_2) \otimes \boldsymbol{w} &= \boldsymbol{v}_2 \otimes \boldsymbol{w} + \lambda (\boldsymbol{v}_2 \otimes \boldsymbol{w}_2).\end{aligned}$$

If $\{\boldsymbol{e}_i\}_{i=1,\ldots,n}$ and $\{\boldsymbol{f}_a\}_{a=1,\ldots,m}$ are bases for V, W respectively, then we have a natural basis:

$$\{\boldsymbol{e}_i \otimes \boldsymbol{f}_a\}_{i=1,\ldots,n;\ a=1,\ldots,m}$$

for $V \otimes W$.

The space $T^p{}_q(V)$

From a real $n-$dimensional vector space V, we can naturally form a $(p+q)n-$dimensional vector space $T^p{}_q(V)$ by taking the tensor product of p copies of V and q copies of V^*:

$$T^p{}_q(V) = \underbrace{V \otimes \cdots \otimes V}_{p \text{ copies}} \otimes \underbrace{V^* \otimes \cdots \otimes V^*}_{q \text{ copies}}.$$

An element of $T^p{}_q(V)$ is called a $(p, q)-$tensor, or a tensor of rank (p, q). A basis $B := \{\boldsymbol{e}_n\}_{i=1,\ldots n}$ induces a basis on $T^p{}_q(V)$:

$$B^p{}_q = \left\{\boldsymbol{e}_{i_1} \otimes \cdots \otimes \boldsymbol{e}_{i_p} \otimes \boldsymbol{e}^{j_1} \otimes \cdots \otimes \boldsymbol{e}^{j_q}\right\}_{i_k, j_l = 1,\ldots,n}.$$

With respect to this basis, we can write any $(p, q)-$tensor $\boldsymbol{T} \in T^p{}_q(V)$ in terms of its components:

$$\boldsymbol{T} = \sum_{i_1,\ldots,i_p,j_1,\ldots,j_q=1}^{n} T^{i_1\ldots i_p}{}_{j_1\ldots j_q} \boldsymbol{e}_{i_1} \otimes \cdots \otimes \boldsymbol{e}_{i_p} \otimes \boldsymbol{e}^{j_1} \otimes \cdots \otimes \boldsymbol{e}^{j_q}.$$

Often the distinction between a tensor and its components is elided, so one will speak of 'the tensor T_{ij}'.

A.1.3 Change of basis

Suppose we have a basis $B := \{\boldsymbol{e}_i\}_{i=1,\ldots,n}$ and we wish to instead work with a new basis, say $B' := \{\boldsymbol{e}'_i\}_{i=1,\ldots,n}$. By virtue of the fact that B' is a basis, we must be able to write

$$\boldsymbol{e}_i = \sum_{j=1}^{n} \boldsymbol{e}'_j \Lambda^j{}_i, \tag{A.1}$$

for some real numbers $\Lambda^j{}_i$ with $i, j = 1, \ldots n$. Equally, we can write

$$\boldsymbol{e}'_i = \sum_{k=1}^{n} \boldsymbol{e}_k \widetilde{\Lambda}^k{}_i, \tag{A.2}$$

for some other real numbers $\widetilde{\Lambda}^j{}_i$ with $i, j = 1, \ldots n$. Substituting (A.1) into (A.2) we find

$$\boldsymbol{e}_i = \sum_{j=1}^{n} \left(\sum_{k=1}^{n} \boldsymbol{e}_k \widetilde{\Lambda}^k{}_j \right) \Lambda^j{}_i = \sum_{k,j=1}^{n} \boldsymbol{e}_k \widetilde{\Lambda}^k{}_j \Lambda^j{}_i$$

Since $\boldsymbol{e}_i$ are linearly independent, we conclude that

$$\sum_{j=1}^{n} \widetilde{\Lambda}^k{}_j \Lambda^j{}_i = \delta^k{}_i.$$

A similar calculation inserting (A.2) into (A.1) shows that

$$\sum_{j=1}^{n} \Lambda^k{}_j \widetilde{\Lambda}^j{}_i = \delta^k{}_i,$$

so that thinking of $\Lambda^i{}_j$ and $\widetilde{\Lambda}^i{}_j$ as the components of a matrix, we have $\Lambda^{-1} = \widetilde{\Lambda}$. In this way, we can identify the set of basis transformations with the group $GL(n, \mathbb{R})$ of invertible linear transformations (or equivalently matrices) on $\mathbb{R}^n$.

Suppose that $\boldsymbol{v} \in V$ has components v^i with respect to the basis B, and components v'^i with respect to the basis B'. We can relate these two sets of components by

$$\boldsymbol{v} = \sum_{i} \boldsymbol{e}_i v^i = \sum_{i,j=1}^{n} \boldsymbol{e}'_j \Lambda^j{}_i v^i = \sum_{i=1}^{n} \boldsymbol{e}'_j v'^j.$$

Using the linear independence of B', we deduce

$$v'^j = \sum_{i=1}^{n} \Lambda^j{}_i v^i, \tag{A.3}$$

so that the components of v change by matrix multiplication on the left by Λ (thinking of v^i as a column vector).

Now let's look at how the dual bases transform. Associated to B and B' are the dual bases $B^* = \{\boldsymbol{e}^i\}_{i=1,\dots n}$ and $B'^* = \{\boldsymbol{e}'^i\}_{i=1,\dots n}$ defined by

$$\boldsymbol{e}^i(\boldsymbol{e}_j) = \delta^i{}_j, \qquad \boldsymbol{e}'^i(\boldsymbol{e}'_j) = \delta^i{}_j, \qquad i,j = 1,\dots,n.$$

I claim that the dual bases are related by the relation:

$$\boldsymbol{e}'^i = \sum_{j=1}^n \Lambda^i{}_j \boldsymbol{e}^j.$$

To show this, it is enough to check that $\boldsymbol{e}'^i(\boldsymbol{e}'_j) = \delta^i{}_j$. We have:

$$\begin{aligned}
\boldsymbol{e}'^i(\boldsymbol{e}'_j) &= \sum_{j=1}^n \Lambda^i{}_k \boldsymbol{e}^k \left(\sum_{k=1}^n \boldsymbol{e}_l \widetilde{\Lambda}^l{}_j \right), \\
&= \sum_{k,l=1}^n \Lambda^i{}_k \widetilde{\Lambda}^l{}_j \boldsymbol{e}^k(\boldsymbol{e}_l), \\
&= \sum_{k=1}^n \Lambda^i{}_k \widetilde{\Lambda}^k{}_j = \delta^i{}_j.
\end{aligned}$$

We can invert the relationship between the bases using the fact that $\widetilde{\Lambda}$ is the matrix inverse of Λ, to find:

$$\boldsymbol{e}^i = \sum_{j=1}^n \widetilde{\Lambda}^i{}_j \boldsymbol{e}'^j.$$

Suppose that $\boldsymbol{\omega} \in V^*$ has components ω_i with respect to the basis B^*, and components $\omega'{}_i$ with respect to the basis B'^*. We can relate these two sets of components by

$$\boldsymbol{\omega} = \sum_i \omega_i \boldsymbol{e}^i = \sum_{i,j=1}^n \omega_i \widetilde{\Lambda}^i{}_j \boldsymbol{e}'^j = \sum_{i=1}^n \omega'_j \boldsymbol{e}'^j$$

Using the linear independence of B'^*, we deduce

$$\omega'{}_j = \sum_{i=1}^n \omega_i \widetilde{\Lambda}^i{}_j, \tag{A.4}$$

so that the components of v change by matrix multiplication on the right by $\widetilde{\Lambda}$ (thinking of ω_i as a row vector).

Extending these arguments to the space $T^p{}_q(V)$, we deduce

Lemma A.1. *Under the change of basis (A.1), the components of a tensor $\boldsymbol{T} \in T^p{}_q(V)$ transform according to:*

$$T'^{i_1\dots i_p}{}_{j_1\dots j_q} = \sum_{k_1,\dots,k_p,l_1,\dots,l_q=1}^n T^{k_1\dots k_p}{}_{l_1\dots l_q} \Lambda^{i_1}{}_{k_1} \cdots \Lambda^{i_p}{}_{k_p} \widetilde{\Lambda}^{l_1}{}_{j_1} \cdots \widetilde{\Lambda}^{l_q}{}_{j_q}.$$

We say that an upstairs index is covariant and a downstairs index is contravariant, reflecting the different transformation laws for the two types of index under a change of basis B for V. For each space $T^p{}_q(V)$, the transformation law gives a *representation* of the group $GL(n, \mathbb{R})$.

Exercise A.1. a) Suppose that a tensor $\boldsymbol{T} \in T^1{}_1(V)$ has components $T^i{}_j = \lambda \delta^i{}_j$ with respect to a given basis B. Show that $\boldsymbol{T}$ has the same components with respect to any other basis. Deduce that the Kronecker delta is invariant under a change of coordinates.

b) Suppose that $\boldsymbol{v} \in V$, with components v^i and $\omega \in V^*$ with components ω_j. Show that the numbers

$$T^i{}_j := v^i \omega_j,$$

transform as the components of a $(1,1)-$tensor. In this way we can build tensors of higher rank from lower rank tensors.

c) Suppose that $\boldsymbol{T} \in T^0{}_2(V)$ has components T_{ij}. Show that the numbers

$$\overline{T}_{ij} := T_{ji},$$

transform as a $(0,2)-$tensor. Deduce that

$$T_{(ij)} := \frac{1}{2}\left(T_{ij} + T_{ji}\right), \quad \text{and} \quad T_{[ij]} := \frac{1}{2}\left(T_{ij} - T_{ji}\right),$$

transform as the components of $(0,2)-$tensors. We call the tensors with components $T_{(ij)}$ and $T_{[ij]}$ the symmetric and antisymmetric part of $\boldsymbol{T}$ respectively.

Contracting indices

Let's suppose we have a tensor $\boldsymbol{T} \in T^1{}_2(V)$, so that its components can be written $T^i{}_{jk}$. Let us consider the following set of numbers:

$$S_k := \sum_{j=1}^{n} T^j{}_{jk}.$$

Now, let us see how S_k transforms when we change our basis. We have

$$T'^i{}_{jk} = \sum_{p,q,r=1,\ldots,n} T^p{}_{qr} \Lambda^i{}_p \widetilde{\Lambda}^q{}_j \widetilde{\Lambda}^r{}_k,$$

so that

$$\begin{aligned} S'_k := \sum_{j=1}^{n} T'^j{}_{jk} &= \sum_{j,p,q,r=1}^{n} T^p{}_{qr} \Lambda^j{}_p \widetilde{\Lambda}^q{}_j \widetilde{\Lambda}^r{}_k, \\ &= \sum_{p,q,r=1}^{n} T^p{}_{qr} \delta^q{}_p \widetilde{\Lambda}^r{}_k = \sum_{pr=1}^{n} T^p{}_{pr} \widetilde{\Lambda}^r{}_k, \\ &= \sum_{r=1}^{n} S_r \widetilde{\Lambda}^r{}_k. \end{aligned}$$

In other words, the set of numbers S_k transforms in the same way as the components of a covector. This means that the map

$$\sum_{i,j,k=1}^{n} T^i{}_{jk}\boldsymbol{e}_i \otimes \boldsymbol{e}^j \otimes \boldsymbol{e}^k \mapsto \sum_{i,j=1}^{n} T^i{}_{ij}\boldsymbol{e}^j,$$

is defined *independently of the basis* with respect to which we express the tensors. We can calculate the sum over components in any basis that we like, and we will still have the same covector as a result. This map is known as a *contraction over indices.* More generally we have

Lemma A.2. *Suppose* $\boldsymbol{T} \in T^p{}_q(V)$ *has components* $T^{i_1 \cdots i_p}{}_{j_1 \cdots j_q}$ *with respect to some basis* B *for* V*. The numbers*[2]

$$T^{i_1 \ldots, \hat{i}_r, \ldots, i_p}{}_{j_1 \ldots, \hat{j}_s, \ldots, j_q} = \sum_{i_r, j_s=1}^{n} T^{i_1 \ldots i_p}{}_{j_1 \ldots, j_q}\delta^{i_r}{}_{j_s},$$

transform in the same fashion as the components of a $(p-1, q-1)-$*tensor. As a result, this defines a natural map* $T^p{}_q(V) \to T^{p-1}{}_{q-1}(V)$ *which does not depend on the choice of basis. This map is called 'contraction of the* r*'th covariant and* s*'th contravariant indices'.*

Summation convention

If we examine the various formulae that appear in this section, we will notice certain patterns. In particular:

- Whenever an index appears exactly once on the left hand side of an equation, the same index appears exactly once on the right hand side, and this index is *not* summed over. These indices are called 'free indices'.

- Whenever an index appears exactly twice, it occurs once in the upstairs position and once in the downstairs position and is summed over from 1 to n. These indices are called 'dummy indices'.

- No index appears more than twice.

These features are the basis for a very powerful notation known as Einstein's summation convention. The basic idea is very simple: we follow the rules as stated above, and we simply leave out any summation signs. Since the summations only appear when we see a repeated index pair with one up and one down, we can always put them back in if we want to. The great power of this convention is that so long as we stick to the rules, the objects we form will always have predictable transformation rules under a change of

[2] $\hat{i}_r$ means omit this index

basis[3]. For example, if $\boldsymbol{T} \in T^3{}_1(V)$ and $\boldsymbol{S} \in T^0{}_2(V)$, then

$$P_l = T^{jk}{}_l S_{jk} = \sum_{j,k=1}^{n} T^{jk}{}_l S_{jk},$$

transform as the components of a covector.

A.1.4 Metric tensors

When looking at vectors on $\mathbb{R}^n$, a useful concept is that of the inner product (sometimes called the dot product). If we take B to be the canonical basis for $\mathbb{R}^n$, then the inner product of two vectors $\boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^n$ is

$$\boldsymbol{v} \cdot \boldsymbol{w} = \sum_{i=1}^{n} v^i w^i.$$

It would be nice to write this using summation convention, however we have the problem that both indices are upstairs, so we can't simply drop the sum. To fix this, we introduce a new $(0,2)-$tensor, $\boldsymbol{g}$, called the *metric tensor*. We define $\boldsymbol{g}$ to have components with respect to the canonical basis:

$$g_{ij} = \begin{cases} 1 & i = j, \\ 0 & i \neq j. \end{cases}$$

Notice that under a change of basis g_{ij} will change in general. This is in contrast to $\delta^i{}_j$ which is the same with respect to any basis. They are different tensors, even though in this basis it looks like they have the same components: the position of indices matters! Having defined $\boldsymbol{g}$, we can write the inner product of two vectors as

$$\boldsymbol{v} \cdot \boldsymbol{w} = g_{ij} v^i w^j.$$

More generally, we will say that a *metric tensor* is a symmetric, non-degenerate, $(0,2)-$tensor. A $(0,2)-$tensor is symmetric if $g_{ij} = g_{ji}$ and non-degenerate if

$$g_{ij} v^i w^j = 0 \text{ for all } w^j \implies v^i = 0.$$

This condition is equivalent to requiring the matrix with components g_{ij} to be invertible. We denote the components of the inverse of g_{ij} by g^{ij}. These satisfy

$$g^{ij} g_{jk} = g_{kj} g^{ji} = \delta^i{}_k. \tag{A.5}$$

Exercise A.2. Verify that if the numbers g^{ij} are the unique solution of (A.5) then they indeed transform as the components of a $(2,0)-$tensor. This tensor is called the cometric.

[3] In this regard, summation convention is similar to Newspeak:

> "In the end we shall make thoughtcrime literally impossible, because there will be no words in which to express it."
>
> *1984, George Orwell*

The metric tensor allows us to identify V and V^* as follows. An element $\boldsymbol{v} \in V$ with components v^i is identified with the element $\boldsymbol{v}^\flat$ with components

$$v_i := g_{ij} v^j.$$

Similarly, an element $\boldsymbol{\omega} \in V^*$ with components ω_i is identified with the element $\boldsymbol{\omega}^\sharp \in V$ with components

$$\omega^i = g^{ij} \omega_j.$$

The notation $\sharp$, $\flat$ is inspired by musical notation and is designed to recall 'raising' and 'lowering' an index. The identification between V and V^* induced by $\boldsymbol{g}$ is occasionally called, somewhat whimsically, the *musical isometry.* This obviously extends to higher rank tensors, and allows us to identify elements of ${T^p}_q(V)$ with elements of ${T^{p'}}_{q'}(V)$ so long as $p + q = p' + q'$. By convention, when writing the components of two tensors identified via $\boldsymbol{g}$, we use the same core letter, relying on the raised and lowered indices to indicate which space the tensor belongs to. When we have a metric tensor, we will often speak of a rank k–tensor, without specifying the index structure.

A.1.5 The orthogonal groups

Given a metric tensor, $\boldsymbol{g}$, we naturally have a bilinear form on V, defined by

$$\boldsymbol{g}(\boldsymbol{v}, \boldsymbol{w}) = g_{ij} v^i w^i = v^i w_i = v_i w^i.$$

By a Gram-Schmidt type of process, it is possible to construct a basis $B = \{\boldsymbol{e}_i\}_{i=1,\ldots,n}$ such that for some $r \in \{1, \ldots, n+1\}$ we have

$$\boldsymbol{g}(\boldsymbol{e}_i, \boldsymbol{e}_j) = g_{ij} = \begin{cases} -1 & i = j, \quad j < r, \\ 1 & i = j, \quad j \geq r, \\ 0 & i \neq j. \end{cases} \tag{A.6}$$

Exercise A.3. Adapt the Gram-Schmidt process to show that there always exists a basis such that (A.6) holds.

We call such a basis, B, an orthonormal basis. With respect to an orthonormal basis, the metric tensor $\boldsymbol{g}$ (thought of as a matrix) is diagonal, with entries ± 1. The number of positive and negative signs is fixed, independent of which orthonormal basis we choose, by Silvester's law of inertia. The number of positive and negative signs is known as the *signature* of the metric, and is usually written as $(+, +, +)$ or $(-, -, +, +)$, or alternatively as (r, s), where r is the number of negative signs and s the number of positive. There are two cases that are most often studied. If all entries are positive, we say the metric has Riemannian signature, and in this case it defines a positive-definite inner product on V. If we have one negative entry and the rest positive, or one positive and the rest negative, we say the metric has Lorentzian signature. This is the case of interest for special and general relativity.

Exercise A.4. Consider $\mathbb{R}^3$ with its canonical basis. Find an orthonormal basis for the metric whose components are given by

$$g_{12} = g_{21} = g_{33} = 1, \qquad \text{all other components } 0.$$

and write down the signature of the metric.

Having established that every metric admits at least one orthonormal basis, a natural question arises concerning other orthonormal bases. Suppose we have a basis B with respect to which $\boldsymbol{g}$ has components given by (A.6). Let B' be a new basis which is also orthonormal, and suppose $\Lambda^i{}_j$ is the matrix corresponding to this change of basis. Then by the transformation law for components of $T^0{}_2(V)$, we have that

$$g'_{ij} = \widetilde{\Lambda}^k{}_i \widetilde{\Lambda}^l{}_j g_{kl},$$

so that if $g'_{ij} = g_{ij}$, we deduce:

$$g_{ij} = \widetilde{\Lambda}^k{}_i \widetilde{\Lambda}^l{}_j g_{kl}. \tag{A.7}$$

Exercise(∗). Show that (A.7) holds if and only if:

$$g_{ij} = \Lambda^k{}_i \Lambda^l{}_j g_{kl}, \tag{A.8}$$

If $\Lambda^i{}_j$ satisfies either (A.7) or (A.8), we say that the corresponding transformation is *orthogonal.* Orthogonal transformations form a group, which can be identified with a subgroup of $GL(n, \mathbb{R})$. We denote this group by $O(r, s)$, where (r, s) is the signature of the metric tensor $\boldsymbol{g}$. In the case that $r = 0$, we have $O(0, n) \equiv O(n)$, the standard orthogonal group. To see this, we note that

$$\Lambda^k{}_i \Lambda^l{}_j \delta_{kl} = \left(\Lambda^T \Lambda\right)_{ij}.$$

The group at the heart of Special Relativity is $O(1, 3)$, which is sufficiently important that it has its own name, the Lorentz group.

A.2 Review of differentiation in $\mathbb{R}^n$

We will review some material about differentiation, and fix some notation that will (hopefully) be familiar if you attended the Manifolds course.

Suppose U is an open subset of $\mathbb{R}^n$ and suppose we have a function $f \in C^1(U)$. This implies that at each point in $\boldsymbol{x}$ there exists a linear map $df|_{\boldsymbol{x}} : \mathbb{R}^n \to \mathbb{R}$ such that if $\boldsymbol{V} \in \mathbb{R}^n$ is any vector, and s is sufficiently small, we have

$$f(\boldsymbol{x} + s\boldsymbol{V}) = f(\boldsymbol{x}) + s\ df|_{\boldsymbol{x}} \boldsymbol{V} + o(s).$$

The linear map $df|_{\boldsymbol{x}}$ is called the *differential* of f at $\boldsymbol{x}$. Geometrically, we should think of the vector $\boldsymbol{V}$ in this formula as having its base at $\boldsymbol{x}$. We call the space of such vectors $T_{\boldsymbol{x}}U$, which is isomorphic, as a vector space, with $\mathbb{R}^3$. Since $df|_{\boldsymbol{x}}$ is a linear map from $T_{\boldsymbol{x}}U$ to $\mathbb{R}$, it belongs to the dual space of $T_{\boldsymbol{x}}U$, which we denote $T^*_{\boldsymbol{x}}U$. Given any vector $\boldsymbol{V} \in T_{\boldsymbol{x}}U$, we can define the *directional derivative* of f along $\boldsymbol{V}$ at $\boldsymbol{x}$ to be:

$$\boldsymbol{V}[f](\boldsymbol{x}) := df|_{\boldsymbol{x}} \boldsymbol{V} = \lim_{s \to 0} \frac{f(\boldsymbol{x} + s\boldsymbol{V}) - f(\boldsymbol{x})}{s}.$$

A.2.1 Working in coordinates

Now, suppose $\{\boldsymbol{e}_i\}_{i=1,\ldots,n}$ is a basis for $\mathbb{R}^n$, and let $\{\boldsymbol{e}^j\}_{i=1,\ldots,n}$ be the dual basis for $(\mathbb{R}^n)^*$, defined by

$$\boldsymbol{e}^i(\boldsymbol{e}_j) = \delta^i{}_j.$$

We define the functions

$$\begin{array}{rcccl} x^i & : & \mathbb{R}^3 & \to & \mathbb{R} \\ & & \boldsymbol{x} & \mapsto & \boldsymbol{e}^i(\boldsymbol{x}) \end{array}$$

for $i = 1, \ldots, n$. It follows[4] that we can write:

$$\boldsymbol{x} = x^i \boldsymbol{e}_i. \tag{A.9}$$

For any vector $\boldsymbol{V} \in T^*_{\boldsymbol{x}} U$, we have

$$\left. dx^i \right|_{\boldsymbol{x}} (\boldsymbol{V}) = \boldsymbol{V}[x^i](\boldsymbol{x}) = \lim_{s\to 0} \frac{\boldsymbol{e}^i(\boldsymbol{x} + s\boldsymbol{V}) - \boldsymbol{e}^i(\boldsymbol{x})}{s} = \boldsymbol{e}^i(\boldsymbol{V}).$$

In other words, we have $dx^i = \boldsymbol{e}^i$.

Now consider a function f which is continuously differentiable in a neighbourhood of $\boldsymbol{x} \in U$, and let us look at $\boldsymbol{V}[f](\boldsymbol{x})$, the directional derivative of f along $\boldsymbol{V}$. We can write $\boldsymbol{V} = V^i \boldsymbol{e}_i$, so that

$$\boldsymbol{V}[f](\boldsymbol{x}) = \left. df \right|_{\boldsymbol{x}} (V^i \boldsymbol{e}_i) = V^i \left. df \right|_{\boldsymbol{x}} (\boldsymbol{e}_i) = V^i \boldsymbol{e}_i[f](\boldsymbol{x}). \tag{A.10}$$

In other words, to calculate the directional derivative along an arbitrary direction, it is enough to know the three directional derivatives $\boldsymbol{e}_i[f](\boldsymbol{x})$. These directional derivatives are useful enough that we give them a special symbol. We write

$$\frac{\partial}{\partial x^i} f(\boldsymbol{x}) := \boldsymbol{e}_i[f](\boldsymbol{x}) = \left. \frac{d}{ds} f(\boldsymbol{x} + s\boldsymbol{e}_i) \right|_{s=0}.$$

This of course agrees with our usual notion of partial derivative. We can go further though, and declare that the object $\frac{\partial}{\partial x^i}$ is itself a vector, which we can identify with $\boldsymbol{e}_i$ i.e. we have:

$$\frac{\partial}{\partial x^i} = \boldsymbol{e}_i \in T_{\boldsymbol{x}} U.$$

We will often find it useful to write as shorthand:

$$\frac{\partial}{\partial x^i} := \partial_i$$

Of course, we have:

$$dx^i(\partial_j) = \delta^i{}_j.$$

Returning to (A.10), we write

$$\boldsymbol{V}[f](\boldsymbol{x}) = V^i \boldsymbol{e}_i[f](\boldsymbol{x}) = \boldsymbol{e}^i(\boldsymbol{V}) \boldsymbol{e}_i[f](\boldsymbol{x}) = \left. \frac{\partial f}{\partial x^i} dx^i \right|_{\boldsymbol{x}} (\boldsymbol{V})$$

[4]Check this. You may need to refer to §A.1

so that we can write

$$df|_{\boldsymbol{x}} = \left. \frac{\partial f}{\partial x^i} dx^i \right|_{\boldsymbol{x}}.$$

Because we have respected the Einstein summation convention, the right hand side of this formula is independent of the choice of basis that we originally made (as, of course it must be).

Exercise A.5. a) Suppose that we consider a new basis $\{\boldsymbol{e}'_i\}$ for $\mathbb{R}^3$. Show that the set of numbers

$$\frac{\partial f}{\partial x^i},$$

transform in the same way as the components of a *co-vector*.

b) Fix $\boldsymbol{x} \in U$ and suppose that $\boldsymbol{V}, \boldsymbol{W} \in T_{\boldsymbol{x}}U$ are two vectors such that for any $f \in C^1(U)$ we have

$$\boldsymbol{V}[f](\boldsymbol{x}) = \boldsymbol{W}[f](\boldsymbol{x}).$$

Show that $\boldsymbol{V} = \boldsymbol{W}$.

A.3 Differential geometry

We will briefly review the basic definitions of a manifold, the tangent bundle and higher rank tensor bundles.

A.3.1 Manifolds

Definition 21. A C^k-*atlas* of a second countable, Hausdorff, topological space $\mathcal{M}$ is a collection of *charts* $\{(\mathcal{U}_\alpha, \phi_\alpha)\}$ which satisfy:

i) Each $\mathcal{U}_\alpha \subset \mathcal{M}$ is an open subset of $\mathcal{M}$, and the $\mathcal{U}_\alpha$ cover $\mathcal{M}$.

ii) φ_α is a homeomorphism from $\mathcal{U}_\alpha$ onto an open subset of $\mathbb{R}^n$.

iii) If $\mathcal{U}_\alpha \cap \mathcal{U}_\beta \neq \emptyset$, then

$$\varphi_{\alpha\beta} := \varphi_\beta \circ \varphi_\alpha^{-1} : \varphi_\alpha\left(\mathcal{U}_\alpha \cap \mathcal{U}_\beta\right) \to \varphi_\beta\left(\mathcal{U}_\alpha \cap \mathcal{U}_\beta\right)$$

is a C^k-diffeomorphism between subsets of $\mathbb{R}^n$, that is to say that $\varphi_{\alpha\beta}^{-1}$ exists and both $\varphi_{\alpha\beta}$ and $\varphi_{\alpha\beta}^{-1}$ are C^k-functions on their respective domains. The functions $\varphi_{\alpha\beta}$ are called *transition functions.*

We say that two C^k-atlases $\{(\mathcal{U}_\alpha, \varphi_\alpha)\}$ and $\{(\mathcal{V}_\alpha, \psi_\alpha)\}$ for $\mathcal{M}$ are compatible if their union is again a C^k-atlas. Clearly compatibility defines an equivalence relation.

Definition 22. A C^k-manifold is a second countable, Hausdorff, topological space $\mathcal{M}$, equipped with an equivalence class of C^k-atlases. The dimension of the manifold is n, where we understand the charts as mapping into $\mathbb{R}^n$.

We can define in an exactly analogous way, a smooth (C^∞) or real analytic[5] (C^ω) atlas (and hence manifold) by requiring that the transition functions be smooth or real analytic diffeomorphisms between open subsets of $\mathbb{R}^n$. We will always assume that $k \geq 1$, which means that objects such as the tangent space are well defined.

Obviously if we have a C^k-manifold, we can always extract from it a $C^{k'}$-manifold for $k' < k$ by picking a representative C^k-atlas and considering it's equivalence class among $C^{k'}$-atlases. There is a pitfall for the unwary in that these two manifolds, although based on the same topological space, are *not* the same. It is, however, common practice to leave the regularity of the manifold unstated in many circumstances so one should be careful.

Example 14. Let $U \subset \mathbb{R}^n$ be an open set with the subset topology. This is certainly a second countable, Hausdorff, topological space. We can equip U with a trivial C^ω-atlas, given by $\{(U, id)\}$. Thus U, with the equivalence class of atlases defined by the trivial atlas, is a real analytic manifold.

Example 15. We define S^1 to be the second countable, Hausdorff[6], topological space $\mathbb{R}/_{\sim_\circ}$, where

$$x \sim_\circ y \quad \Longleftrightarrow \exists\, n \in \mathbb{Z} \text{ s.t. } \quad x = y + 2\pi n.$$

We can define an atlas as follows. Let $0 \leq \alpha < 2\pi$. We set $\mathcal{U}_\alpha = S^1 \setminus [\alpha]_{\sim_\circ}$, and define

$$\begin{array}{rrcl} \varphi_\alpha \ : & \mathcal{U}_\alpha & \to & (0, 2\pi) \\ & [x]_{\sim_\circ} & \mapsto & \theta - \alpha \end{array}$$

where θ is the unique real number satisfying $\theta \sim_\circ x$ and $\alpha < \theta < \alpha + 2\pi$. Suppose $0 \leq \alpha < \alpha' < 2\pi$. We have that

$$\begin{aligned} \varphi_\alpha\left(\mathcal{U}_\alpha \cap \mathcal{U}_\beta\right) &= (0, 2\pi) \setminus \{\alpha' - \alpha\} \\ \varphi_\beta\left(\mathcal{U}_\alpha \cap \mathcal{U}_\beta\right) &= (0, 2\pi) \setminus \{2\pi + \alpha - \alpha'\} \end{aligned}$$

and

$$\varphi_{\alpha\beta}(s) = \begin{cases} s + 2\pi + \alpha - \alpha' & 0 < s < \alpha' - \alpha, \\ s - \alpha' + \alpha & \alpha' - \alpha < s < 2\pi. \end{cases}$$

This can be easily verified to be a real analytic diffeomorphism, so that $\mathscr{A} = \{\mathcal{U}_\alpha, \varphi_\alpha\}_{\alpha \in [0, 2\pi)}$ is a real analytic atlas for S^1. Taking $\mathscr{A}$ to define an equivalence class of C^ω-atlases, we can make S^1 into a real analytic manifold.

A.3.2 Mappings between manifolds, and their derivatives

Smooth functions

Suppose that we have two manifolds $\mathcal{M}, \mathcal{N}$, which are both at least C^k-regular[7] and are equipped with representative atlases $\{(U_\alpha, \varphi_\alpha)\}, \{(\mathcal{V}_\beta, \psi_\beta)\}$. Consider a function:

$$f : \mathcal{M} \to \mathcal{N}.$$

[5]A function f is real analytic if there is a neighbourhood of every point on which f can be expressed as a convergent Taylor series.

[6]you should check that this is true

[7]From now on, we'll abuse notation and allow '$k = \infty$' and '$k = \omega$' in statements like this.

We say that f is C^k-smooth if for any α, β, the function

$$\psi_\beta \circ f \circ \varphi_\alpha^{-1} : \mathcal{U}_\alpha \to \mathcal{V}_\beta$$

is C^k, understood in the usual way for maps between subsets of $\mathbb{R}^n$ and $\mathbb{R}^m$. We denote by $C^k(\mathcal{M};\mathcal{N})$ the set of all C^k-smooth maps from $\mathcal{M}$ to $\mathcal{N}$.

Exercise(∗). Show that this definition is independent of the choice of atlases. That is, if we choose a different atlas $\{(U'_\alpha, \varphi'_\alpha)\}$ for $\mathcal{M}$ which is compatible with $\{(U_\alpha, \varphi_\alpha)\}$, our definition of $C^k(\mathcal{M};\mathcal{N})$ agrees for both.

Exercise A.6. a) Show that the identity map on a C^k-manifold, $\mathcal{M}$, always belongs to $C^k(\mathcal{M};\mathcal{M})$.

b) Show that if $f, g \in C^k(\mathcal{M};\mathbb{R})$, we have $fg \in C^k(\mathcal{M};\mathbb{R})$, where we define the product pointwise $fg(p) = f(p)g(p)$.

c) Suppose $\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3$ are all at least C^k regular, and that $f \in C^k(\mathcal{M}_1;\mathcal{M}_2)$, $g \in C^k(\mathcal{M}_2;\mathcal{M}_3)$. Show that $g \circ f \in C^k(\mathcal{M}_1;\mathcal{M}_3)$.

d) Let $\mathcal{M}$ be a C^k-manifold of dimension n and let $\pi^i : \mathbb{R}^n \to \mathbb{R}$ be the projection onto the i^{th} coordinate. Suppose that $(\mathcal{U}_\alpha, \varphi_\alpha)$ is a chart. Show that

$$\varphi^i_\alpha = \pi^i \circ \varphi_\alpha : \mathcal{U}_\alpha \to \mathbb{R}$$

is C^k-smooth.

Example 16. Take S^1 and $\mathbb{R}^2$ with their real analytic structures as previously defined. We define a map

$$\begin{array}{rccl} f & : & S^1 & \to & \mathbb{R}^2, \\ & & [x]_{\sim_\circ} & \mapsto & (\sin x, \cos x). \end{array}$$

First, note that this is well defined regardless of which representative x for $[x]_{\sim_\circ}$ we choose. Now consider the functions

$$\begin{array}{rccl} id \circ f \circ \varphi_\alpha^{-1} & : & (0, 2\pi) & \to & \mathbb{R}^2, \\ & & \theta & \mapsto & (\sin(\theta + \alpha), \cos(\theta + \alpha)). \end{array}$$

which are clearly real analytic. Thus, $f \in C^\omega(S^1;\mathbb{R}^2)$.

The tangent and co-tangent space at a point

Now that we have defined our C^k-smooth functions, we can define the tangent vectors to the manifold $\mathcal{M}$ at a point $p \in M$. There are various ways of doing this, some more concrete than others. We'll give the definition here, and then show how this definition fits with our intuition from the case of $\mathbb{R}^n$.

We say that a C^k-smooth map $\gamma : (-\epsilon, \epsilon) \to \mathcal{M}$ is a curve in $\mathcal{M}$. Fix $p \in \mathcal{M}$. We define the tangent space at p, $T_p\mathcal{M}$ to be the space of curves in $\mathcal{M}$ with $\gamma(0) = p$, modulo the equivalence relation:

$$\gamma_1 \sim \gamma_2 \quad \text{if} \quad \left.\frac{d}{dt}\left[f \circ \gamma_1\right](t)\right|_{t=0} = \left.\frac{d}{dt}\left[f \circ \gamma_2\right](t)\right|_{t=0} \quad \text{for all } f \in C^k(\mathcal{M};\mathbb{R}).$$

Each element of $T_p\mathcal{M}$ (i.e. each equivalence class of curves under $\sim$) is called a tangent vector at p, and we write $[\gamma]_\sim = \dot{\gamma}(0)$. For any tangent vector $V \in T_p\mathcal{M}$, and function $f \in C^k(\mathcal{M};\mathbb{R})$, we define the directional derivative of f along V at p to be:

$$V[f](p) := \left.\frac{d}{dt}\left[f \circ \gamma\right](t)\right|_{t=0}$$

for any γ such that $\dot{\gamma}(0) = V$. This is a useful way to think of a tangent vector at p. It is a direction in which we can differentiate a function.

We can endow $T_p\mathcal{M}$ with a vector space structure using the following result:

Lemma A.3. *Let $(\mathcal{U}_\alpha, \varphi_\alpha)$ be a chart with $p \in \mathcal{U}_\alpha$. The chart induces a canonical identification between $T_p\mathcal{M}$ and $\mathbb{R}^n$.*

Proof. Let us set $\varphi_\alpha(p) = \boldsymbol{x}$. Suppose we are given two curves γ_1, γ_2 with $\gamma_i(0) = p$. We can use the fact that $\varphi_\alpha^{-1} \circ \varphi_\alpha = id_{\mathcal{U}_\alpha}$ to see that

$$\left.\frac{d}{dt}\left[f \circ \gamma_1\right](t)\right|_{t=0} = \left.\frac{d}{dt}\left[f \circ \gamma_2\right](t)\right|_{t=0} \qquad \text{for all } f \in C^k(\mathcal{M};\mathbb{R})$$

holds if and only if

$$\left.\frac{d}{dt}\left[f \circ \varphi_\alpha^{-1} \circ \varphi_\alpha \circ \gamma_1\right](t)\right|_{t=0} = \left.\frac{d}{dt}\left[f \circ \varphi_\alpha^{-1} \circ \varphi_\alpha \circ \gamma_2\right](t)\right|_{t=0} \qquad \text{for all } f \in C^k(\mathcal{M};\mathbb{R})$$

which is true if and only if

$$\left.\frac{d}{dt}\left[\widetilde{f} \circ \widetilde{\gamma}_1\right](t)\right|_{t=0} = \left.\frac{d}{dt}\left[\widetilde{f} \circ \widetilde{\gamma}_2\right](t)\right|_{t=0} \qquad \text{for all } \widetilde{f} \in C^k\left(\varphi_\alpha(\mathcal{U}_\alpha);\mathbb{R}\right).$$

Where for $i = 1, 2$ we have defined

$$\begin{array}{rcrcl} \widetilde{\gamma}_i & : & (-\epsilon, \epsilon) & \to & \varphi_\alpha(\mathcal{U}_\alpha), \\ & & t & \mapsto & (\varphi_\alpha \circ \gamma_i)(t) \end{array}$$

which is a curve in a subset of $\mathbb{R}^n$. As a result, we can simply apply the chain rule to deduce that $\gamma_1 \sim \gamma_2$ if and only if we have equality of the directional derivatives:

$$\boldsymbol{V}_1[f](\boldsymbol{x}) = \boldsymbol{V}_2[f](\boldsymbol{x}), \qquad \text{for all } f \in C^k(\varphi_\alpha(\mathcal{U}_\alpha);\mathbb{R}),$$

where $\boldsymbol{V}_i = \dot{\widetilde{\gamma}}(0)$ are the tangent vectors to the curves $\widetilde{\gamma}_i$, in the usual sense of curves in $\mathbb{R}^n$. By Exercise A.5 this holds if and only if $\boldsymbol{V}_1 = \boldsymbol{V}_2$. Thus, with each equivalence class $[\gamma]_\sim$ we can associate a unique vector in $\mathbb{R}^n$ via the coordinate chart $(\mathcal{U}_\alpha, \varphi_\alpha)$. Conversely, from $\boldsymbol{V} \in \mathbb{R}^n$, we can construct the curve

$$\gamma_{\boldsymbol{V}}(t) = \varphi_\alpha^{-1}\left(\boldsymbol{x} + \boldsymbol{V}t\right)$$

Which satisfies $\dot{\widetilde{\gamma}}_{\boldsymbol{V}}(0) = \boldsymbol{V}$. Thus, the chart $(\mathcal{U}_\alpha, \varphi_\alpha)$ induces a canonical identification of $T_p\mathcal{M}$ with $\mathbb{R}^n$. ∎

We deduce from this Lemma the important result:

Theorem A.1. *Given $\gamma_i : (-\epsilon, \epsilon) \to \mathcal{M}$ with $\gamma_i(0) = p$ for $i = 1, 2$ and $\lambda \in \mathbb{R}$, there exists a curve γ such that*

$$\frac{d}{dt}\left[f \circ \gamma\right](t)\bigg|_{t=0} = \frac{d}{dt}\left[f \circ \gamma_1\right](t)\bigg|_{t=0} + \lambda\, \frac{d}{dt}\left[f \circ \gamma_2\right](t)\bigg|_{t=0}$$

for all $f \in C^k(\mathcal{M}; \mathbb{R})$. Clearly, this defines $[\gamma]_\sim = \dot{\gamma}(0)$ uniquely. We then write

$$\dot{\gamma}_1(0) + \lambda\dot{\gamma}_2(0) := \dot{\gamma}(0)$$

This provides us with a definition of addition of tangent vectors and scalar multiplication. With these definitions, T_pM is a real vector space of dimension n.

Proof. We may simply take (with the notation of the Lemma)

$$\gamma(t) = \gamma_{\boldsymbol{V}}(t)$$

where

$$\boldsymbol{V} = \dot{\tilde{\gamma}}_1(0) + \lambda\dot{\tilde{\gamma}}_2(0).$$

This gives us a vector space isomorphism between $T_p\mathcal{M}$ and $\mathbb{R}^n$. □

We can use this canonical identification induced by $(\mathcal{U}_\alpha, \varphi_\alpha)$ to provide us with a basis for $T_p\mathcal{M}$ for each $p \in \mathcal{U}_\alpha$. We pick a basis $\{\boldsymbol{e}_i\}_{i=1,\dots,n}$ for $\mathbb{R}^n$ and define

$$\left(\frac{\partial}{\partial x^i}\right)_p = (\partial_i) := \dot{\gamma}_{\boldsymbol{e_i}}(0).$$

We say that $\left\{\left(\frac{\partial}{\partial x^i}\right)_p\right\}$ is a coordinate basis for $T_p\mathcal{M}$.

The reason for this notation is made clear by the following result

Lemma A.4. *Suppose $f \in C^k(\mathcal{M}; \mathbb{R})$ and let $(\mathcal{U}_\alpha, \varphi_\alpha)$ be a chart with $p \in \mathcal{U}_\alpha$ and $\varphi_\alpha(p) = \boldsymbol{x}$. We define $\widetilde{f} = f \circ \varphi_\alpha^{-1}$, which is a C^k function defined on some open set about $\boldsymbol{x} = x^i\boldsymbol{e}_i$. We then have:*

$$\left(\frac{\partial}{\partial x^i}\right)_p [f](p) = \left(\frac{\partial \widetilde{f}}{\partial x^i}\right)(\boldsymbol{x})$$

where on the right hand side, we simply have the partial derivative of $\widetilde{f}$ as a function on $\mathbb{R}^n$.

Proof. We simply calculate using the definitions:

$$\begin{aligned}\left(\frac{\partial}{\partial x^i}\right)_p [f](p) &= \frac{d}{dt}\left[f \circ \gamma_{\boldsymbol{e_i}}\right]\bigg|_{t=0} \\ &= \frac{d}{dt}\left[(f \circ \varphi_\alpha^{-1})(\boldsymbol{x} + \boldsymbol{e_i}t)\right]\bigg|_{t=0} \\ &= \left(\frac{\partial \widetilde{f}}{\partial x^i}\right)(\boldsymbol{x}).\end{aligned}$$

□

Now that we have the vector space $T_p\mathcal{M}$, we are free to construct its dual, and the higher rank tensor spaces associated with it. Particularly important is the dual space $T_p^*\mathcal{M}$, known as the co-tangent space at p. The cotangent space is the natural place in which the differential of a function lives:

Definition 23. Given $f \in C^k(\mathcal{M};\mathbb{R})$, we define the differential of f at p, $df|_p$ to be the linear map

$$\begin{array}{rccc} df|_p & : & T_p\mathcal{M} & \to & \mathbb{R}, \\ & & V & \mapsto & V[f](p). \end{array}$$

We can define a basis for $T_p^*\mathcal{M}$ from the coordinate basis for $T_p\mathcal{M}$ by duality. We define $dx^i\big|_p \in T_p^*\mathcal{M}$ by the requirement that

$$dx^i\big|_p \left[\left(\frac{\partial}{\partial x^j} \right)_p \right] = \delta^i{}_j.$$

With this definition we have

$$df|_p = \left(\frac{\partial}{\partial x^i} \right)_p [f](p)\ \ dx^i\big|_p$$

The tangent and co-tangent bundles

In the previous subsection, we restricted our attention to vectors and co-vectors defined at a single point. More often than not, rather than considering a vector at a single point we want to consider a *vector field*, where we have a vector defined at each point. Before discussing vector fields, we introduce the *tangent bundle*. This is the disjoint union over the tangent spaces at each point in the manifold:

$$T\mathcal{M} = \bigsqcup_{p\in\mathcal{M}} T_p\mathcal{M} = \bigcup_{p\in\mathcal{M}} \{p\} \times T_p\mathcal{M}.$$

We shall show that a C^k-atlas for $\mathcal{M}$ induces a C^{k-1}-atlas for $T\mathcal{M}$, so that the tangent bundle may be given a manifold structure in a natural fashion. Suppose $\{(\mathcal{U}_\alpha, \varphi_\alpha)\}$ is a C^k-atlas for $\mathcal{M}$. Pick a chart $(\mathcal{U}_\alpha, \varphi_\alpha)$. We know that at each point $p \in \mathcal{U}_\alpha$ the chart induces a canonical identification between $T_p\mathcal{M}$ and $\mathbb{R}^n$. A tangent vector $V \in T_p\mathcal{M}$ is identified uniquely with a vector $\boldsymbol{V} \in \mathbb{R}^n$. We now define:

$$\hat{\mathcal{U}}_\alpha = \bigsqcup_{p\in\mathcal{U}_\alpha} T_p\mathcal{M}$$

and

$$\begin{array}{rccc} \hat{\varphi}_\alpha & : & \hat{\mathcal{U}}_\alpha & \to & \varphi_\alpha(\mathcal{U}_\alpha) \times \mathbb{R}^n, \\ & & (p, V) & \mapsto & (\varphi_\alpha(p), \boldsymbol{V}). \end{array}$$

Clearly $\hat{\varphi}_\alpha$ maps $\hat{\mathcal{U}}_\alpha$ bijectively onto an open subset of $\mathbb{R}^{2n}$. We define a topology on $T\mathcal{M}$ by declaring that $\mathcal{A} \subset T\mathcal{M}$ is open if and only if $\hat{\varphi}_\alpha\left(\mathcal{A} \cap \hat{\mathcal{U}}_\alpha\right)$ is open in $\mathbb{R}^{2n}$ for any α. One can verify that this definition turns $T\mathcal{M}$ into a second countable, Hausdorff, topological space.

Lemma A.5. *The collection* $\{(\hat{\mathcal{U}}_\alpha, \hat{\varphi}_\alpha)\}$ *is a* C^{k-1}*-atlas for* $T\mathcal{M}$.

Proof. By construction, $\hat{\varphi}_\alpha$ maps $\hat{\mathcal{U}}_\alpha$ homeomorphically onto an open subset of $\mathbb{R}^{2n}$ and moreover the sets $\hat{\mathcal{U}}_\alpha$ cover $T\mathcal{M}$. It remains to show that the transition functions are C^{k-1}-diffeomorphisms between subsets of $\mathbb{R}^{2n}$. I claim that we have

$$\begin{array}{rcrcl} \hat{\varphi}_{\alpha\beta} & : & \varphi_\alpha(\mathcal{U}_\alpha \cap \mathcal{U}_\alpha) \times \mathbb{R}^n & \to & \varphi_\beta(\mathcal{U}_\alpha \cap \mathcal{U}_\alpha) \times \mathbb{R}^n, \\ & & (\boldsymbol{x}, \boldsymbol{V}_\alpha) & \mapsto & (\varphi_{\alpha\beta}(\boldsymbol{x}), D\varphi_{\alpha\beta}|_{\boldsymbol{x}} (\boldsymbol{V}_\alpha)). \end{array}$$

To see this, recall that if $\varphi_\alpha(p) = \boldsymbol{x}$, then $\boldsymbol{V}_\alpha \in \mathbb{R}^n$ corresponds to $V \in T_p\mathcal{M}$ under the identification induced by φ_α implies that $V = \dot{\gamma}(0)$ for the curve

$$\begin{array}{rcccc} \gamma & : & (-\epsilon, \epsilon) & \to & \mathcal{M}, \\ & & t & \mapsto & \varphi_\alpha^{-1}(\boldsymbol{x} + t\boldsymbol{V}_\alpha). \end{array}$$

Recall also that, going in the other direction, if $V \in T_p\mathcal{M}$, then under the identification induced by φ_β it corresponds to the vector $\boldsymbol{V}_\beta \in \mathbb{R}^n$ defined by

$$\boldsymbol{V}_\beta = \frac{d}{dt}(\varphi_\beta \circ \gamma)(t)\bigg|_{t=0}$$

for any $\gamma : (-\epsilon, \epsilon) \to \mathcal{M}$ such that $\gamma(0) = p$ and $\dot{\gamma}(0) = V$. Combining these two facts, we have that

$$\begin{aligned} \boldsymbol{V}_\beta &= \frac{d}{dt}(\varphi_\beta \circ \varphi_\alpha^{-1})(\boldsymbol{x} + t\boldsymbol{V}_\alpha)\bigg|_{t=0} \\ &= \frac{d}{dt}(\varphi_{\alpha\beta})(\boldsymbol{x} + t\boldsymbol{V}_\alpha)\bigg|_{t=0} \\ &= D\varphi_{\alpha\beta}|_{\boldsymbol{x}} (\boldsymbol{V}_\alpha). \end{aligned}$$

Now, since $\varphi_{\alpha\beta}$ is a C^k-diffeomorphism, $D\varphi_{\alpha\beta}$ is a C^{k-1} map into the space of invertible matrices, thus the transition function $\hat{\varphi}_{\alpha\beta}$ is a C^{k-1}-diffeomorphism. □

As a consequence, we have:

Theorem A.2. *The tangent bundle naturally inherits the structure of a* C^{k-1}*-manifold of dimension* $2n$.

An entirely analogous construction can be performed in which we glue together the co-tangent spaces to define the co-tangent bundle, $T^*\mathcal{M}$. This again inherits the structure of a C^{k-1}-manifold of dimension $2n$ from $\mathcal{M}$. The only difference in our development above is that the relevant transition functions are given by[8]:

$$\begin{array}{rcrcl} \hat{\varphi}^*_{\alpha\beta} & : & \varphi_\alpha(\mathcal{U}_\alpha \cap \mathcal{U}_\alpha) \times (\mathbb{R}^n)^* & \to & \varphi_\beta(\mathcal{U}_\alpha \cap \mathcal{U}_\alpha) \times (\mathbb{R}^n)^*, \\ & & (\boldsymbol{x}, \boldsymbol{\omega}_\alpha) & \mapsto & (\varphi_{\alpha\beta}(\boldsymbol{x}), D\varphi_{\alpha\beta}^{-1}\big|_{\boldsymbol{x}}^* (\boldsymbol{\omega}_\alpha)). \end{array}$$

[8] recall that the adjoint of a linear map $\Lambda : V \to V$ is a linear map $\Lambda^* : V^* \to V^*$ whose action on $\omega \in V^*$ is given by:

$$(\Lambda^* \omega)[v] = \omega[\Lambda v], \qquad \text{for all } v \in V.$$

Glueing together tensor products of $T_p\mathcal{M}$ and $T_p^*\mathcal{M}$ in the same manner, we can construct tensor bundles of arbitrary rank.

Definition 24. Suppose $\mathcal{M}$ is at least $k+1$ regular.

i) A C^k-vector field is a C^k-map $X : \mathcal{M} \to T\mathcal{M}$ such that at every point $p \in \mathcal{M}$, we have $X(p) \in T_p(\mathcal{M})$. The set of all C^k-vector fields is denoted $\mathfrak{X}(\mathcal{M})$, or $\mathfrak{X}_k(\mathcal{M})$ if the regularity is not obvious from context.

ii) A C^k-one-form is a C^k-map $\omega : \mathcal{M} \to T^*\mathcal{M}$ such that at every point $p \in \mathcal{M}$, we have $\omega(p) \in T_p^*(\mathcal{M})$. The set of all C^k-one-form fields is denoted $\mathfrak{X}^*(\mathcal{M})$, or $\mathfrak{X}_k^*(\mathcal{M})$ if the regularity is not obvious from context.

Notice that there is a natural paring $\mathfrak{X}_k(\mathcal{M}) \times \mathfrak{X}_k^*(\mathcal{M}) \to C^k(\mathcal{M};\mathbb{R})$ defined by

$$p \in \mathcal{M} \mapsto \omega|_p \left(X|_p \right).$$

We can extend the definition to higher rank tensor fields in the obvious fashion.

Suppose that $\mathcal{M}$ is a C^k-manifold and that $\varphi : \mathcal{U} \subset \mathcal{M} \to U \subset \mathbb{R}^n$ is a coordinate chart, and pick a basis $\{\boldsymbol{e}_i\}$ for $\mathbb{R}^n$. At each point, we have a basis for $T_p\mathcal{M}$ given by

$$\left\{ \left(\frac{\partial}{\partial x^i} \right)_p \right\}.$$

For each i, the map

$$\begin{array}{rccc} \frac{\partial}{\partial x^i} & : \mathcal{U} & \to & T\mathcal{U} \subset T\mathcal{M} \\ & p & \mapsto & \left(\frac{\partial}{\partial x^i}\right)_p \end{array}$$

defines a $C^{k-1}-$vector field on $\mathcal{U}$, i.e.

$$\frac{\partial}{\partial x^i} \in \mathfrak{X}_{k-1}(\mathcal{U}).$$

This vector field is also sometimes written ∂_i. In a similar fashion, dual to these vector fields we have the one-forms

$$dx^i \in \mathfrak{X}_{k-1}^*(\mathcal{U})$$

satisfying

$$dx^i \left(\frac{\partial}{\partial x^j} \right) = \delta^i{}_j.$$

We know that at each point $p \in \mathcal{U}$ that $\{(\partial_i)_p\}_{i=1,\ldots,n}$, $\{ dx^i\big|_p\}_{i=1,\ldots,n}$ span $T_p\mathcal{M}$, $T_p^*\mathcal{M}$ respectively. As a consequence, any C^r-smooth $(p,q)-$tensor field with $r \le k-1$ can be written locally as

$$T = T^{\mu_1,\ldots\mu_p}{}_{\nu_1,\ldots\nu_q} \frac{\partial}{\partial x^{\mu_1}} \otimes \cdots \otimes \frac{\partial}{\partial x^{\mu_p}} \otimes dx^{\nu_1} \otimes \cdots \otimes dx^{\mu_q}$$

Where $T^{\mu_1,\ldots\mu_p}{}_{\nu_1,\ldots\nu_q} : \mathcal{U} \to \mathbb{R}$ are C^r-functions, called the components of T in the coordinate chart $(\varphi,\mathcal{U})$. Using the chain rule, it is straightforward to demonstrate that the condition that the components are C^r-smooth is independent of the coordinate chart, provided $r \le k-1$.

A.3.3 Derivations and commutators

Definition 25. A C^r-*derivation* on the C^k-manifold $\mathcal{M}$ is an $\mathbb{R}-$linear map

$$D : C^{s+1}(\mathcal{M}; R) \to C^s(\mathcal{M}; R)$$

for any $s \leq r$, which satisfies

i) $D(f_1 + \lambda f_2) = Df_1 + \lambda D f_2$,

ii) $D(f_1 f_2) = f_1 D f_2 + f_2 D f_1$,

for all $f_i \in C^{s+1}(\mathcal{M}; R)$, $\lambda \in \mathbb{R}$.

A vector field $V \in \mathfrak{X}_r(\mathcal{M})$ naturally defines a C^r-derivation. We can check that

$$D_V f(p) := V_p[f] = \frac{d}{dt} f \circ \gamma(t)\bigg|_{t=0}$$

where $\gamma(0) = p, \dot{\gamma}(0) = V_p$ satisfies the conditions above. Conversely, we can identify a C^r-derivation with a C^r-vector field. To see this, we first note that any C^1-function in $\mathbb{R}^n$ may be written locally about a point y as

$$g(x) = g(y) + g_i(x)(x^i - y^i),$$

where $g_i \in C^0(\mathbb{R}^n)$ with $g_i(y) = \frac{\partial g}{\partial x^i}(y)$. Now consider a point $p \in \mathcal{M}$ and a coordinate chart $(\mathcal{U}, \varphi)$ with $p \in \mathcal{U}$ and $\varphi = (\varphi^1, \ldots, \varphi^n)$. We deduce that any function $f \in C^1(\mathcal{M}; \mathbb{R})$ may be written for q in a neighbourhood of p as:

$$f(q) = f(p) + f_i(q)(\varphi^i(q) - \varphi^i(p))$$

for some $f_i \in C^0(\mathcal{M}; \mathbb{R})$ with $f_i(p) = \frac{\partial}{\partial x^i}\big|_p [f]$. Applying an arbitrary derivation to this formula, we deduce that

$$Df = (\varphi^i - \varphi^i(p))Df_i + f_i D\varphi^i$$

Evaluating this formula at p we deduce:

$$Df(p) = f_i(p) D\varphi^i(p) = \frac{\partial}{\partial x^i}\bigg|_p [f] D\varphi^i(p) = V|_p [f]$$

where $V|_p \in T_p\mathcal{M}$ is the tangent vector given in local coordinates by:

$$V|_p = D\varphi^i(p) \left.\frac{\partial}{\partial x^i}\right|_p$$

Defining $V \in \mathfrak{X}_r(\mathcal{U})$ by

$$V := D\varphi^i \frac{\partial}{\partial x^i}$$

we have

$$D = D_V.$$

We have shown that in a neighbourhood of any point there exists a V such that $D = D_V$. Since a vector field is uniquely determined by its action on functions, V is uniquely fixed by this condition.

Lemma A.6. *Let $\mathcal{M}$ be a C^k manifold. Suppose $X, Y \in \mathfrak{X}_r(\mathcal{M})$ for $r < k$. There exists a unique vector field $[X, Y] \in \mathfrak{X}_{r-1}(\mathcal{M})$ defined by the condition:*

$$[X, Y]f = X(Yf) - Y(Xf), \qquad \forall f \in C^r(\mathcal{M}; \mathbb{R})$$

Proof. By the previous discussion, it suffices to check that the operation $[X, Y]f$ defines a $C^{r-1}-$derivation. We calculate

$$\begin{aligned} [X, Y](f_1 + \lambda f_2) &= X(Y(f_1 + \lambda f_2)) - Y(X(f_1 + \lambda f_2)) \\ &= X(Yf_1) - Y(Xf_1) + \lambda\left[X(Yf_2) - Y(Xf_2)\right] \\ &= [X, Y]f_1 + \lambda[X, Y]f_2, \end{aligned}$$

using the $\mathbb{R}-$linearity of the action of a vector field on a function. We also find

$$\begin{aligned} [X, Y](fg) &= X(Y(fg)) - Y(X(fg)) \\ &= X(fYg + gYf) - Y(fXg + gXf) \\ &= fX(Yg) + (Xf)(Yg) + (Xg)(Yf) + gX(Yf) \\ &\quad - \left[fY(Xg) + (Yf)(Xg) + (Yg)(Xf) + gY(Xf)\right] \\ &= f\left[X(Yg) - Y(Xg)\right] + g\left[X(Yf) - Y(Xf)\right] \\ &= f[X, Y]g + g[X, Y]f \end{aligned}$$

Hence the operation $f \mapsto [X, Y]f$ is a derivation, to which we can associate the unique vector field $[X, Y]$. □

Exercise(∗). Working in a coordinate patch, $\mathcal{U}$ so that we can write

$$X = X^i \frac{\partial}{\partial x^i}, \qquad X = Y^i \frac{\partial}{\partial x^i},$$

for $X^i, Y^i \in C^r(\mathcal{U}; \mathbb{R})$, show that

$$[X, Y] = \left(X^i \frac{\partial Y^j}{\partial x^i} - Y^i \frac{\partial X^j}{\partial x^i}\right) \frac{\partial}{\partial x^j}.$$

A.3.4 Immersions and embeddings

Push forward, pull back

For this section, we shall make the assumption that all manifolds and objects that we discuss are smooth. We return to the situation we considered earlier, where we have two smooth manifolds $\mathcal{N}, \mathcal{M}$, of dimension $n, n + d$ respectively. Suppose we have a function $\phi \in C^\infty(\mathcal{N}; \mathcal{M})$. The mapping ϕ naturally induces relations between various geometric objects defined on the two manifolds $\mathcal{N}$, $\mathcal{M}$. We start with the pull back of a function by ϕ. Suppose $f \in C^\infty(\mathcal{M}; \mathbb{R})$, we define the *pull back* of f by ϕ, written $\phi^* f \in C^\infty(\mathcal{N}; \mathbb{R})$ by

$$\phi^* f := f \circ \phi.$$

Now let us consider vectors. Suppose $X \in T_p\mathcal{N}$, and that $\gamma \in C^\infty((-\epsilon,\epsilon),\mathcal{N})$ satisfies $\gamma(0) = p$, $\dot{\gamma}(0) = X$. We can define a curve, $\widetilde{\gamma} \in C^\infty((-\epsilon,\epsilon),\mathcal{M})$ by:

$$\widetilde{\gamma} := \phi \circ \gamma$$

We define the *push-forward* of X by ϕ, written $\phi_* X \in T_{\phi(p)}\mathcal{M}$ by:

$$\phi_* X = \dot{\widetilde{\gamma}}(0).$$

Lemma A.7. *Suppose that $X \in \mathfrak{X}(\mathcal{N})$, $f \in C^\infty(\mathcal{M};\mathbb{R})$ and $\phi \in C^\infty(\mathcal{N},\mathcal{M})$. Fix $p \in \mathcal{N}$. Then:*

$$X(\phi^* f)|_p = [(\phi_* X)\, f]|_{\phi(p)}\,,$$

and

$$f \phi_* X|_{\phi(p)} = \phi_* \left[(\phi^* f) X\right]|_{\phi(p)}$$

Proof. Fix $p \in \mathcal{N}$. Suppose that γ satisfies $\gamma(0) = p$, $\dot{\gamma}(0) = X|_p$. From the definition of a vector acting on a function, we have:

$$\begin{aligned}
X(\phi^* f)|_p &= \left.\frac{d}{ds}(\phi^* f) \circ \gamma(s)\right|_{s=0} \\
&= \left.\frac{d}{ds}(f \circ \phi) \circ \gamma(s)\right|_{s=0} \\
&= \left.\frac{d}{ds} f \circ (\phi \circ \gamma)(s)\right|_{s=0} \\
&= \left.\frac{d}{ds} f \circ \widetilde{\gamma}(s)\right|_{s=0} \\
&= (\phi_* X)\, f|_{\phi(p)}
\end{aligned}$$

For the second part, suppose that $g \in C^\infty((M);\mathbb{R})$, and calculate using the previous result:

$$\begin{aligned}
[f\,(\phi_* X)\, g]|_{\phi(p)} &= f(\phi(p)) X(\phi^* g)|_p \\
&= (\phi^* f)\,(p)\; X(\phi^* g)|_p \\
&= [(\phi^* f)\, X(\phi^* g)]|_p \\
&= \phi_* \left[(\phi^* f) X\right] g|_{\phi(p)}
\end{aligned}$$

□

Now, suppose that $\omega \in T^*_{\phi(p)}\mathcal{M}$ is a one-form. We can define the *pull-back* of ω by ϕ, written $\phi^*\omega \in T_p\mathcal{N}$ by:

$$\left[\phi^*\omega\right](X) = \omega\left(\phi_* X\right), \qquad \forall\; X \in T_p\mathcal{N}.$$

These definitions readily extend to allow us to define the push forward of any $(p,0)-$tensor, and the pull-back of any $(0,q)-$tensor. Notice that the push forward and pull-back at each point p are linear maps between vector spaces.

Exercise(*). Suppose $f \in C^\infty(\mathcal{M};\mathbb{R})$ and $\phi \in C^\infty(\mathcal{N},\mathcal{M})$. Show that:

$$d\left(\phi^* f\right) = \phi^* df.$$

Notation: Where the choice of map ϕ is clear from context, we will write $f^* = \phi^* f$, $X_* = \phi_* X$, etc.

Definition 26. We say that a map $\imath \in C^\infty(\mathcal{N};\mathcal{M})$ is a smooth *immersion* if the push forward map:

$$\imath_* : T_p\mathcal{N} \to T_{\phi(p)}\mathcal{M}$$

is an injective linear map for each $p \in \mathcal{N}$. We say that $\imath$ is a smooth *embedding* if moreover $\imath$ is injective. In this case, we write $\imath : \mathcal{N} \hookrightarrow \mathcal{M}$.

If we have that $\imath \in C^\infty(\mathcal{N};\mathcal{M})$ is an immersion, then for each $p \in \mathcal{N}$, we can identify $T_p\mathcal{N} \simeq T_{\imath(p)}\imath(\mathcal{N})$, where $T_{\imath(p)}\imath(\mathcal{N})$ is a linear subspace of $T_{\phi(p)}\mathcal{M}$.

Canonical immersions and extensions

Lemma A.8 (Canonical Immersion Theorem)**.** *Suppose $\imath : \mathcal{N} \hookrightarrow \mathcal{M}$ is an immersion, and fix $p \in \mathcal{N}$. There exist coordinate charts $(\mathcal{U},\varphi)$ and $(\mathcal{V},\psi)$ for $\mathcal{N}$ and $\mathcal{M}$ respectively, with $\varphi(p) = 0$, such that*

$$\begin{array}{rcrcl} \psi\circ\imath\circ\varphi^{-1} & : & \varphi(\mathcal{U}) & \to & \psi(\mathcal{V}), \\ & & (x_1,\ldots,x_n) & \mapsto & (x_1,\ldots,x_n,0,\ldots,0). \end{array}$$

Proof. Let us fix $(\mathcal{U},\varphi)$ to be a coordinate chart for $\mathcal{N}$ centred at p, i.e., $\varphi(p) = 0$. Now consider any $(\mathcal{V},\psi)$, a coordinate chart for $\mathcal{M}$ centred at $\imath(p)$, and set $f = \psi\circ\imath\circ\varphi^{-1}$. We will denote by $(x_i)_{i=1,\ldots n}$ the coordinates on $\varphi(\mathcal{U}) \subset \mathbb{R}^n$, and by $(y_a)_{a=1,\ldots,n+d}$ the coordinates of $\psi(\mathcal{V}) \subset \mathbb{R}^{n+d}$.

The fact that $\imath$ is an immersion implies that $Df(0)$ is injective, so by a linear transformation on $\mathbb{R}^{n+d}$, we may assume

$$Df(0) = \left(\begin{array}{cc} I_{n\times n} & 0_{d\times n} \end{array}\right).$$

Now set

$$h(y_1,\ldots,y_{n+d}) = f(y_1,\ldots y_n) + (0,\ldots,0,y_{n+1},\ldots,y_{n+d}),$$

defined on some neighbourhood of 0 in $\mathbb{R}^{d+n}$. Clearly $h(0) = 0$ and $Dh(0) = I_{(n+d)\times(n+d)}$, so by restricting the domain to a subset if necessary, we can assume that h is smoothly invertible. We can easily verify that on a sufficiently small neighbourhood of 0, the map $h^{-1}\circ f$ takes $(x_1,\ldots,x_n)$ to $(x_1,\ldots,x_n,0,\ldots,0)$, so by redefining ψ to be $\psi\circ h$ and shrinking $\mathcal{U},\mathcal{V}$ as necessary, we are done. □

From here, we can derive various extension Lemmas which allow us to locally extend objects defined on $\imath(\mathcal{N})$ to a neighbourhood in $\mathcal{M}$ in a smooth fashion. Obviously such extensions are highly non-unique.

Corollary A.3. *Suppose $\imath : \mathcal{N} \to \mathcal{M}$ is an immersion, and let $p \in \mathcal{N}$. There exists a neighbourhood $\mathcal{W}$ of p in M such that for any $f \in C^\infty(\mathcal{N};\mathbb{R})$ we can find a function $\widetilde{f} \in C^\infty(\mathcal{M};\mathbb{R})$ with $f = \imath^* \widetilde{f}$ in $\mathcal{W}$.*

Proof. For $p \in \mathcal{N}$ we have local coordinate charts $(\mathcal{U}, \varphi)$, $(\mathcal{V}, \psi)$ centred at $p, \imath(p)$ such that $\imath$ is given by the canonical immersion. Pick $W \Subset \varphi(\mathcal{U})$, and define $\mathcal{W} = \varphi^{-1}(W)$. Let $\chi_T : U \to [0,1]$ be a smooth cut-off function equal to unity on W and vanishing near the boundary of $\varphi(\mathcal{U})$. We define $\overline{f} = f \circ \varphi^{-1}$ and finally we set $\widetilde{f} = 0$ on $\mathcal{M} \setminus \mathcal{V}$, and $\widetilde{f} = \overline{\overline{f}} \circ \psi$, for

$$\overline{\overline{f}}(y_1, \ldots y_{n+d}) = \chi_N(y_{n+1}, \ldots, y_{n+d})\chi_T(y_1, \ldots, y_n)\overline{f}(y_1, \ldots, y_n).$$

Here $\chi_N : \mathbb{R}^d \to [0,1]$ is a smooth cut-off function, equal to 1 at the origin and supported on a sufficiently small set that

$$\text{supp}\ (\chi_N(y_{n+1}, \ldots, y_{n+d})\chi_T(y_1, \ldots, y_n)) \subset \psi(\mathcal{V}).$$

□

Let $\mathcal{U} \subset \mathcal{N}$ be an open set, and $X \in \mathfrak{X}(\mathcal{U})$. We say that $\widetilde{X} \in \mathfrak{X}(\mathcal{M})$ is an extension of X away from $\imath(\mathcal{U})$ if $X(f^*) = \imath^* \left[\widetilde{X}(f)\right]$ in $\mathcal{U}$ for any $\phi \in C^\infty(\mathcal{M};\mathbb{R})$. Equivalently, $X_* = \widetilde{X}$ on $\imath(\mathcal{U})$.

Corollary A.4. *Suppose $\imath : \mathcal{N} \to \mathcal{M}$ is an immersion and let $p \in \mathcal{N}$. There exists a neighbourhood $\mathcal{W}$ of p in $\mathcal{N}$ such that for any $X \in \mathfrak{X}(\mathcal{N})$ we can find $\widetilde{X} \in \mathfrak{X}(\mathcal{M})$ which is an extension of X away from $\imath(\mathcal{W})$.*

Proof. Take the same charts as in the previous proof, so that (x^i) are local coordinates on $\mathcal{N}$ and (y^a) are local coordinates on $\mathcal{M}$. We have that:

$$X = \sum_{i=1}^n X^i \frac{\partial}{\partial x^i},$$

for $X^i \in C^\infty(\mathcal{U}, \mathbb{R})$. By the previous result, there exists $\mathcal{W} \subset \mathcal{N}$ containing p, and $\widetilde{X}^i \in C^\infty(\mathcal{N};\mathbb{R})$ such that $\imath^* \widetilde{X}^i = X^i$. Defining $\widetilde{X} \in \mathfrak{X}(\mathcal{M})$ by:

$$\widetilde{X} = \sum_{i=1}^n \widetilde{X}^i \frac{\partial}{\partial y^i}$$

we have the desired extension away from $\imath(\mathcal{W})$. □

This allows us to prove the following result

Lemma A.9. *Suppose $X, Y \in \mathfrak{X}(\mathcal{N})$, and pick $p \in \mathcal{N}$. Let $\widetilde{X}, \widetilde{Y} \in \mathfrak{X}(\mathcal{M})$ be extensions of X, Y away from $\imath(\mathcal{U})$ for some neighbourhood $\mathcal{U}$ of p. Then $\left[\widetilde{X}, \widetilde{Y}\right]$ is an extension of $[X, Y]$ away from $\imath(\mathcal{U})$.*

Proof. We calculate in $\mathcal{U}$:

$$\begin{aligned}\left(\left[\widetilde{X},\widetilde{Y}\right](\phi)\right)^* &= \left[\widetilde{X}\left(\widetilde{Y}(\phi)\right)\right]^* - \left[\widetilde{Y}\left(\widetilde{X}(\phi)\right)\right]^* \\ &= X\left(\left[\widetilde{Y}(\phi)\right]^*\right) - Y\left(\left[\widetilde{X}(\phi)\right]^*\right) \\ &= X\left(Y\left(\phi^*\right)\right) - Y\left(X\left(\phi^*\right)\right) \\ &= [X,Y]\left(\phi^*\right)\end{aligned}$$

whence we are done. □

Diffeomorphisms

We say that the map $\phi \in C^\infty(\mathcal{N};\mathcal{M})$ is a *diffeomorphism* if ϕ^{-1} exists and belongs to $C^\infty(\mathcal{M};\mathcal{N})$. A diffeomorphism allows us to identify $T_p\mathcal{N} \simeq T_{\phi(p)}\mathcal{M}$ for all $p \in \mathcal{N}$.

We can construct a family of diffeomorphisms taking $\mathcal{M}$ to itself from a vector field $X \in \mathfrak{X}(\mathcal{M})$. For this, we require the following result:

Lemma A.10. *Let $X \in \mathfrak{X}(\mathcal{M})$, and suppose $p \in \mathcal{M}$. Then there exists a parameterised curve $\gamma \in C^\infty((a,b);\mathcal{M})$, where $a < 0 < b$, satisfying:*

$$\gamma(0) = p, \qquad \dot{\gamma}(s) = X|_{\gamma(s)}\,.$$

This is called an integral curve *of X starting at p, and is unique up to extension.*

Proof. Take a coordinate chart $(\varphi,\mathcal{U})$ with $p \in \mathcal{U}$, and write $X = X^i \frac{\partial}{\partial x^i}$, for $X^i \in C^k(\mathcal{U};\mathbb{R})$, and set $\varphi \circ \gamma(s) = x(s) = (x^i(s))$. The condition on γ becomes:

$$\dot{x}^i(s) = X^i \circ \varphi^{-1}(x(s)), \qquad x(0) = \varphi(p),$$

which is an ordinary differential equation, with Lipshitz right hand side, whose solutions are unique up to extension. By this local uniqueness property, the curve is defined independent of the coordinate chart, and is unique up to extension. □

Lemma A.11. *Suppose that $X \in \mathfrak{X}(\mathcal{M})$ has the property that for each $p \in \mathcal{M}$, the integral curve of X through p, written γ_p can be extended so that it belongs to $C^k((-\epsilon,\epsilon);\mathcal{M})$ for some ϵ independent of p. Then the map*

$${}^X\phi_s : p \mapsto \gamma_p(s)$$

is a diffeomorphism, referred to as the one parameter family of diffeomorphisms induced by X.

A.4 Matrix Lie Groups

This section is meant to give a very brief introduction to Lie groups. We bypass a lot of important theory and focus our attention on the matrix Lie groups. As a result, some of the definitions are not standard, but allow us to get to our goal more quickly. It is certainly no substitute for a proper study of the beautiful topic of Lie groups. Think of the approach here as 'good enough' for our purposes.

We'll start by extending the familiar definition of a group to encompass an additional requirement, that of differentiability.

Definition 27. A *group* is a set G, together with a binary operation $\cdot$ such that

i) $a \cdot b \in G$ for all $a, b \in G$ [Closure]

ii) $(a \cdot b) \cdot c = a \cdot (b \cdot c)$ for all $a, b \in G$ [Associativity]

iii) There exits $e \in G$ such that $e \cdot a = a \cdot e = a$ for all $a \in G$ [Existence of an identity]

iv) For each $a \in G$ there exists $b \in G$ such that $a \cdot b = b \cdot a = e$ [Existence of an inverse]

The identity element is unique, as is the inverse, and we write $b = a^{-1}$ if $a \cdot b = b \cdot a = e$.

A *Lie group* is a group, where the set G has a smooth[9] structure making it into a manifold in such a way that the operations $\cdot$ and $()^{-1}$ are smooth.

The many of the most important Lie groups are groups of matrices:

Example 17. Consider the set $GL(n)$ of invertible $n \times n$ real matrices. This is naturally a group with the group operation of matrix multiplication. Thinking of $GL(n)$ as an open subset of $\mathbb{R}^{n^2}$ with the standard real analytic structure, $GL(n)$ is also a manifold in a natural way. Matrix multiplication and inversion are smooth with respect to this structure, as can be seen by writing out the operations in components.

In fact, this example (and its subgroups) will be sufficiently rich to cover the situations that we are interested in for this course. We define in the obvious way

Definition 28. A *Lie subgroup* of a Lie group G is a subgroup H of G endowed with a topology and smooth structure making it into a Lie group in such a way that the inclusion map is an immersion. A *matrix Lie group* is an embedded Lie subgroup of $GL(n)$.

An important (and deep result) tells us when a subgroup of a Lie group is in fact a Lie subgroup:

Proposition 2 (Closed subgroup theorem)**.** *If H is a subgroup of G which is closed (in the topology of G), then H is an embedded Lie subgroup of G.*

This gives us the following result

Lemma A.12. *Suppose that H is a subgroup of $GL(n)$ with the property that for any sequence $A_n \in H$ with $A_n \to A \in Mat(n \times n)$ component-wise, either $A \in H$ or $A \notin GL(n)$. Then H is a matrix Lie group.*

Example 18.

1. $GL^+(n) = \{A \in GL(n) : \det A > 0\}$ is a matrix Lie group
2. $SL(n) = \{A \in GL(n) : \det A = 1\}$ is a matrix Lie group

[9] It turns out that we can assume that the manifold is C^ω without any loss of generality.

3. $O(n) = \{A \in GL(n) : A^T A = I\}$ is a matrix Lie group

4. The set of matrices $A \in GL(2)$ of the form

$$A = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}, \qquad \theta \in \mathbb{Q},$$

is a subgroup of $GL(2)$, but *not* a matrix Lie group.

A.4.1 The matrix exponential

A very useful tool to understand the structure of matrix Lie groups is the matrix exponential. Before we define the matrix exponential, it's first useful to introduce a norm on $Mat(n \times n)$, called the operator norm. We define

$$||A||_{op.} = \sup_{x \in \mathbb{R}^n, x \neq 0} \frac{||Ax||}{||x||}$$

where $||\cdot||$ is the usual Euclidean norm. We clearly have that for any non-zero $x \in \mathbb{R}^n$:

$$||Ax|| \leq ||A||_{op.} ||x|| .$$

we deduce that

$$||(A+B)x|| \leq ||Ax|| + ||Bx|| \leq ||A||_{op.} ||x|| + ||B||_{op.} ||x||$$

so that

$$||A+B||_{op.} \leq ||A||_{op.} + ||B||_{op.} .$$

so the triangle inequality is satisfied by this norm. The other criteria for $||\cdot||_{op.}$ to define a norm are straightforward to verify.

We can also show by a simple induction that

$$||A^n x|| \leq ||A||^n_{op.} ||x||$$

holds for any x and thus

$$||A^n||_{op.} \leq ||A||^n_{op.} .$$

Definition 29. The matrix exponential of an element $A \in Mat(n \times n)$ is defined to be

$$\text{Exp } A = e^A := \sum_{k=0}^{\infty} \frac{A^n}{n!}. \tag{A.11}$$

Exercise(∗). i) Show that the sum on the right hand side of (A.11) converges in the operator norm for any A (and hence with respect to any other norm on $Mat(n \times n)$). Show also that

$$\left(e^A\right)^T = e^{A^T}$$

ii*) Show that the matrix exponential is a real analytic function $Mat(n \times n) \to GL(n)$, where both spaces inherit the canonical real analytic structure of $\mathbb{R}^{n^2}$, and show that

$$\frac{d}{dt}e^{tA} = Ae^{tA} = e^{tA}A$$

and

$$e^{tA}e^{sA} = e^{(t+s)A} = e^{sA}e^{tA}$$

for $t, s \in \mathbb{R}$. Deduce that

$$\left(e^{A}\right)^{-1} = e^{-A}.$$

iv) Show that

$$\lim_{n\to\infty}\left(1 + \frac{1}{n}A\right)^{n} = e^{A},$$

v) Deduce that if $\Gamma : (-\epsilon, \epsilon) \to GL(n)$ is a C^1-curve with $\Gamma(0) = I$ and $\dot{\Gamma}(0) = A$, we have

$$\lim_{n\to\infty}\left[\Gamma\left(\frac{1}{n}\right)\right]^{n} = e^{A}$$

A.4.2 The Lie algebra

With the matrix exponential in our pocket, we are ready to define the Lie algebra of a matrix Lie group.

Definition 30. The Lie algebra $\mathfrak{h}$ of a matrix Lie group H is defined to be the set of all matrices $a \in Mat(n \times n)$ such that $e^{ta} \in H$ for all $t \in \mathbb{R}$.

Theorem A.5. *Let $\mathfrak{h}$ be the Lie algebra of a matrix Lie group H. Then*

i) $a \in \mathfrak{h} \Leftrightarrow a \in T_I H$, i.e. a is in the Lie algebra if and only if there exists a C^1-curve $\Gamma : (-\epsilon, \epsilon) \to H$ with $\Gamma(0) = I$ and $\dot{\Gamma}(0) = a$.

ii) $\mathfrak{h}$ is a vector subspace of $Mat(n \times n)$, i.e. if $a, b \in \mathfrak{h}$ then $a + \lambda b \in \mathfrak{h}$

iii) $\mathfrak{h}$ is closed under the matrix commutator, i.e. if $a, b \in \mathfrak{h}$ then $[a, b] = ab - ba \in \mathfrak{h}$

Proof. i) "$\Rightarrow$" is trivial, take $\Gamma(t) = e^{ta}$. For "$\Leftarrow$", we use the fact that $\Gamma\left(\frac{t}{n}\right) \in H$ for any $t \in \mathbb{R}$ and n a sufficiently large integer to deduce that

$$\left[\Gamma\left(\frac{t}{n}\right)\right]^{n} \in H$$

for n sufficiently large. We know

$$\lim_{n\to\infty}\left[\Gamma\left(\frac{t}{n}\right)\right]^{n} = e^{ta}$$

which is invertible, so by the closeness of H, we have $e^{ta} \in H$ and we're done.

ii) Suppose $a, b \in \mathfrak{h}$ and fix $\lambda \in \mathbb{R}$. Consider the curve $\Gamma : (-\epsilon, \epsilon) \to H$ given by

$$\Gamma(t) = e^{ta} e^{t\lambda b}$$

This is clearly C^1 and we readily calculate that

$$\Gamma(0) = I, \quad \dot{\Gamma}(0) = a + \lambda b,$$

so that $a + \lambda b \in \mathfrak{h}$.

iii) Now consider the curve $\Gamma : (-\epsilon, \epsilon) \to H$ given by

$$\Gamma(t) = \begin{cases} e^{\sqrt{t}a} e^{\sqrt{t}\lambda b} e^{-\sqrt{t}(a+b)} & t > 0 \\ I & t = 0 \\ e^{\sqrt{-t}b} e^{\sqrt{-t}\lambda a} e^{-\sqrt{-t}(a+b)} & t < 0 \end{cases}$$

This is smooth for $t \neq 0$. Expanding for t small and positive, we have

$$\begin{aligned} \Gamma(t) &= \left[I + \sqrt{t}a + \frac{1}{2}ta^2 + o(t)\right]\left[I + \sqrt{t}b + \frac{1}{2}tb^2 + o(t)\right]\left[I - \sqrt{t}(a+b) + \frac{1}{2}t(a+b)^2 + o(t)\right] \\ &= I + t\left\{\frac{1}{2}\left[a^2 + b^2 + (a+b)^2\right] + ab - a(a+b) - b(a+b)\right\} + o(t) \\ &= I + \frac{t}{2}\left\{ab - ba\right\} + o(t) \end{aligned}$$

Noting that we can obtain the expansion for t small and negative by interchanging $t \leftrightarrow -t$ and $a \leftrightarrow b$, we deduce that Γ is in fact C^1 with $\dot{\Gamma}(0) = \frac{1}{2}[a, b]$.

□

Thus the Lie algebra of a matrix Lie group is naturally a vector space endowed with an antisymmetric bilinear operation $[,]$, which moreover satisfies the Jacobi identity

$$[[a, b], c] + [[b, c], a] + [[c, a], b] = 0$$

One can abstractly define a Lie algebra to be a vector space endowed with such an operation, but the concrete realisation as a space of matrices endowed with the matrix commutator is the most useful for us.

A.4.3 The orthogonal group

As a brief example, we will give a brief treatment of the orthogonal group. Recall

$$O(n) = \{A \in GL(n) : A^T A = I\}.$$

Let us find the Lie algebra, $\mathfrak{o}(n)$. Suppose $\Gamma : (-\epsilon, \epsilon) \to O(n)$ with $\Gamma(0) = I$. We have

$$\Gamma(t)^T \Gamma(t) = I, \quad t \in (-\epsilon, \epsilon)$$

so we can differentiate this condition to find

$$\dot{\Gamma}(t)^T \Gamma(t) + \Gamma(t)^T \dot{\Gamma}(t) = 0, \quad t \in (-\epsilon, \epsilon)$$

so that, setting $t = 0$ we deduce that if $\dot{\Gamma}(0) = a \in \mathfrak{o}(n)$ we must have

$$a^T + a = 0$$

so that a is antisymmetric. Conversely, suppose that a is antisymmetric. We have

$$\left(e^{ta}\right)^T = e^{ta^T} = e^{-ta} = \left(e^{ta}\right)^{-1},$$

so that $e^{ta} \in O(n)$. Thus $\mathfrak{o}(n)$ is precisely the set of antisymmetric matrices.

We should be careful here. Although $e^{ta} \in O(n)$ whenever $a \in \mathfrak{o}(n)$, not every element of $O(n)$ can be written as the exponential of an antisymmetric matrix. We know that if $A \in O(n)$ a simple calculation shows that $\det A = \pm 1$. The determinant of a matrix is a continuous real valued function, so on any continuous curve $\Gamma : (a, b) \to O(n)$ the determinant must be constant. In particular for $a \in \mathfrak{o}(n)$ we have:

$$\det\left(e^{at}\right) = 1.$$

In fact, we can show that by exponentiating elements of $\mathfrak{o}(n)$ it is possible to construct any element of $SO(n)$.

Exercise(∗). Let

$$a_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \quad a_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \quad a_3 = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Show that

$$[a_i, a_j] = \epsilon_{ijk} a_k,$$

where ϵ_{ijk} is the totally antisymmetric tensor with $\epsilon_{123} = 1$. Show also that:

$$e^{\theta a_1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix}$$

and find similar expressions for $e^{\theta a_2}$, $e^{\theta a_3}$. Deduce that by exponentiating $\mathfrak{o}(3)$ we can produce a matrix representing an arbitrary rotation, i.e. an arbitrary element of $S0(3)$.

Appendix B

Bonus material

B.1 Integration in $\mathbb{E}^{1,3}$ and the divergence theorem

We will need to integrate over surfaces embedded in $\mathbb{E}^{1,3}$ and relate these surface integrals to volume integrals using the divergence theorem in the usual way. We would like the expressions we write down to be invariant under a change of frame. This is slightly complicated when the metric has Lorentzian signature. The main issue is that the notion of *outwards unit normal* becomes tricky when vectors can have zero norm.

First, suppose we have an open set $U \subset \mathbb{E}^{1,3}$ and that $f : U \mapsto \mathbb{R}$ is a function defined on U. Picking an inertial frame $\{\vec{e}_\mu\}$, we naturally have that $\widetilde{f} : (x^\mu) \mapsto f(x^\mu \vec{e}_\mu)$ is a map from some subset of $\widetilde{U} \subset \mathbb{R}^4$ to $\mathbb{R}$. We say that f is measurable if $\widetilde{f}$ is Lebesgue measurable. Consider now

$$I[f] := \int_{\widetilde{U}} \widetilde{f}(x) dx$$

where dx is the standard Lebesgue measure on $\mathbb{R}^4$. Now consider a different choice of inertial frame $\{\vec{e}_\mu{}'\}$. We define $\widetilde{f}' : (x'^\mu) \mapsto f(x'^\mu \vec{e}_\mu{}')$, which maps $\widetilde{U}' \subset \mathbb{R}^4$ to $\mathbb{R}$. We have

$$\widetilde{f}(x) = \widetilde{f}'(\Lambda x)$$

Where $\Lambda \in O(1,3)$ is the matrix representing the change of basis. Now, we have

$$I'[f] = \int_{\widetilde{U}'} \widetilde{f}'(x) dx = \int_{\widetilde{U}} \widetilde{f}(y) \left|\det \Lambda\right| dy = \int_{\widetilde{U}} \widetilde{f}(y) dy = I[f]$$

where we change variables $x^\mu = \Lambda^\mu{}_\nu y^\nu$ and use that $|\det \Lambda| = 1$. Thus, we can define

$$\int_U f dX := I[f],$$

and this definition is independent of the inertial frame.

Next, let us introduce the *alternating tensor*. This is a $(0,4)-$tensor, defined to be totally antisymmetric (i.e. antisymmetric under exchange of any pair of indices), and such that in a given inertial frame we have

$$\varepsilon_{0123} = 1.$$

Under a Lorentz transformation with metric $\Lambda^\mu{}_\nu$, the alternating tensor transforms as[1]

$$\varepsilon_{\mu\nu\sigma\tau} = (\det \Lambda)\varepsilon'_{\mu\nu\sigma\tau},$$

so that provided we restrict the transformations to belong to $SO(1,3)$, this tensor is invariant.

Now suppose that we have a surface Σ which may be written as $\Sigma = \varphi(U)$ for a smooth parameterisation φ:

$$\begin{array}{rccc} \vec{\varphi} \; : & U \subset \mathbb{R}^3 & \rightarrow & \Sigma \subset \mathbb{E}^{1,3} \\ & (y^1, y^2, y^3) & \mapsto & \varphi^\mu(y)\boldsymbol{e}_\mu \end{array}$$

The choice of parameterisation naturally gives an orientation to the surface. At any point, the ordered set of vectors

$$\left\{ \frac{\partial \vec{\varphi}}{\partial y^1}, \frac{\partial \vec{\varphi}}{\partial y^2}, \frac{\partial \vec{\varphi}}{\partial y^3} \right\}$$

defines an oriented basis for $T\Sigma$.

The vector surface measure of Σ with respect to this parameterisation is the vector measure (defined on a subset of $\mathbb{R}^3$):

$$dS^{\vec{\varphi}}_\mu = \varepsilon_{\mu\nu\sigma\tau} \frac{\partial \varphi^\nu}{\partial y^1} \frac{\partial \varphi^\sigma}{\partial y^2} \frac{\partial \varphi^\tau}{\partial y^3} dy$$

Where dy is the standard Lebesgue measure on $\mathbb{R}^3$.

For example, suppose we consider the plane $\Sigma = \{t = 0\}$, which we can parameterize by the map:

$$\begin{array}{rccc} \vec{\varphi} \; : & U \subset \mathbb{R}^3 & \rightarrow & \Sigma \subset \mathbb{E}^{1,3} \\ & (y^1, y^2, y^3) & \mapsto & y^i \boldsymbol{e}_i \end{array}$$

so that $\varphi^0(y) = 0, \varphi^i(y) = y^i$. Then we have that for this parameterisation

$$dS^\varphi_0 = dy, \qquad dS^\varphi_i = 0.$$

Exercise(∗). a) Show that if $F : W \to U$ is an orientation preserving diffeomorphism between subsets of $\mathbb{R}^3$ then

$$\int_{y\in U} V^\mu(\vec{\varphi}(y)) dS^{\vec{\varphi}}_\mu = \int_{y \in W} V^\mu(\vec{\varphi} \circ F(y)) dS^{(\vec{\varphi}\circ F)}_\mu.$$

b) Show that if P is the orientation reversing diffeomorphism $P : W \to U$ given by $y \to -y$, then we have

$$\int_{y\in U} V^\mu(\vec{\varphi}(y)) dS^{\vec{\varphi}}_\mu = -\int_{y \in W} V^\mu(\vec{\varphi} \circ P(y)) dS^{(\vec{\varphi}\circ P)}_\mu.$$

Deduce that

$$\int_{y\in U} V^\mu(\vec{\varphi}(y)) dS^{\vec{\varphi}}_\mu$$

depends only on the orientation of Σ and not on the choice of parameterisation.

[1] Check that you understand why this is!

With the vector surface measure, we can define the flux of a sufficiently smooth vector field V through Σ:

$$\int_\Sigma \vec{V} \cdot d\vec{S} := \int_{y \in U} V^\mu(\vec{\varphi}(y)) dS^{\vec{\varphi}}_\mu$$

The flux of $\vec{V}$ through Σ does not depend on the choice of parameterisation, but it does depend on the orientation for the surface Σ. By taking a partition of unity to localise $\vec{V}$ in coordinate patches[2], we can define the flux through any oriented surface[3] $\Sigma \subset \mathbb{E}^{1,3}$.

If Σ is an orientable surface which is everywhere spacelike or timelike, we can define a scalar measure on Σ, which we'll denote by $d\sigma$:

$$\int_\Sigma f d\sigma = \int_\Sigma f \vec{N} \cdot d\vec{S}$$

where $\vec{N}$ is a unit normal of Σ with respect to η (i.e. $\eta(\vec{N}, \vec{N}) = \pm 1$) with the direction chosen such that $N^\mu dS^\varphi_\mu$ is a positive measure on $U \subset \mathbb{R}^3$ for each parameterisation respecting the orientation. Notice that if Σ is not either everywhere spacelike or timelike then this definition does not make sense because a null surface does not have a well defined unit normal. We still have a perfectly reasonable vector measure, but no scalar measure.

Theorem B.1. *Suppose that $\vec{V}$ is a vector field on a bounded domain $U \subset \mathbb{E}^{1,3}$, whose boundary $\Sigma = \partial U$ is piecewise smooth. Suppose also that $V^\mu \in C^1(\overline{U})$. Then we have*

$$\int_U \nabla_\mu V^\mu dX = \int_\Sigma \vec{V} \cdot d\vec{S},$$

where the orientation of Σ is chosen such that if $\vec{K}$ is an outwards pointing vector transverse to Σ, then

$$\{\vec{K}, \vec{V}_1, \vec{V}_2, \vec{V}_3\}$$

is a positively oriented basis for $T_p\mathbb{E}^{1,3}$ whenever $\{\vec{V}_1, \vec{V}_2, \vec{V}_3\}$ is a positively oriented basis for $T_p\Sigma$.

We shall omit the proof here. It can be deduced from the usual divergence theorem on $\mathbb{R}^4$ keeping careful account of changes of sign.

Lemma B.1. *With the same set-up as in Theorem B.1, the choice of orientation for Σ is equivalent to the requirement that*

$$K^\mu dS^\varphi_\mu$$

is a positive measure on $U \subset \mathbb{R}^3$, whenever $\vec{\varphi} : U \to \Sigma$ is a local parameterisation respecting the orientation.

[2]This is a technical point, which can usually be avoided. If the surface Σ cannot be smoothly parameterised by a single coordinate patch then we have to write our vector field $\vec{V}$ as a sum of terms each of which is supported on a single patch. We can then define the integral of $\vec{V}$ over the surface by linearity.

[3]i.e. a surface with a consistent global choice of orientation

Corollary B.2. *Suppose that $\vec{V}$ is a vector field on a bounded domain $U \subset \mathbb{E}^{1,3}$, whose boundary $\Sigma = \partial U$ is piecewise smooth and can be written as $\Sigma = \Sigma_s \cup \Sigma_t$ where Σ_s is spacelike and Σ_t is timelike. Suppose also that $V^\mu \in C^1(\overline{U})$. Then we have*

$$\int_U \nabla_\mu V^\mu dX = \int_{\Sigma_s} \vec{V} \cdot \vec{N} d\sigma - \int_{\Sigma_t} \vec{V} \cdot \vec{N} d\sigma,$$

where $\vec{N}$ is the unit outwards *normal.*

Notice here that there is a sign change for the timelike surfaces relative to the spacelike surfaces.

Example 19. Fix an inertial frame for $\mathbb{E}^{1,3}$. Consider the cylinder $\mathcal{C} = \{x^\mu : 0 < x^0 < 1, |\boldsymbol{x}| < 1\}$. We have:

$$\int_{\mathcal{C}} \nabla_\mu V^\mu dX = \int_{x^0=1, \boldsymbol{x} \in B_1(0)} V^t dx - \int_{x^0=1, \boldsymbol{x} \in B_1(0)} V^t dx + \int_{[0,1] \times \partial B_1(0)} \boldsymbol{V} \cdot \boldsymbol{n} \, d\sigma dt$$

where $B_r(\boldsymbol{x})$ is the Euclidean ball of radius r centred at $\boldsymbol{x}$ and $\boldsymbol{n}, d\sigma$ are the usual outward directed normal and surface measure of the unit Euclidean sphere.

Index