Phys 514

General Relativity

April Sada Solomon

Winter 2021

Course Description

Notes taken directly from the lectures given by Dr. Maloney (not proof read by him). Special Relativity, Manifolds, Spacetime Curvature, Gravitation, Schwarzchild Solution, Black Holes, Perturbations, Radiation, Introduction to Cosmology, Introduction to QFT in Spacetime.

Contents

1 A Brief Review of Special Relativity

1.1 Introduction

General Relativity is the modern theory of classical gravity. What do we mean by "classical"? Simply, it is a nomenclature to distinguish from Quantum Mechanics, so those effects are ignored. The basic idea from Newtonian Mechanics is that gravity is a force, and as we know, most forces are described by fields.

The simplest example of a field that describes a force is the Newtonian gravitational potential $\Phi(\vec{r})$:

$$\frac{d}{dt}\frac{\partial \mathscr{L}}{\partial \dot{x}} = \frac{\partial \mathscr{L}}{\partial x} \implies \vec{F}(\vec{v}) = -\frac{\partial}{\partial \vec{r}}\Phi(\vec{r}) \tag{1.1}$$

If we are studying electromagnetism, we would study the electromagnetic fields $\vec{E}$ and $\vec{B}$. Let us list the Maxwell Equations, that we will review again later on:

$$\nabla \cdot \vec{E} = \frac{\rho}{\varepsilon_0} \tag{1.2}$$

$$\nabla \cdot \vec{B} = 0 \tag{1.3}$$

$$\nabla \times \vec{E} = -\frac{\partial \vec{B}}{\partial t} \tag{1.4}$$

$$\nabla \times \vec{B} = \mu_0 \left(\vec{J} + \varepsilon_0 \frac{\partial \vec{E}}{\partial t}\right) \tag{1.5}$$

A classical field theory like the theory of Newtonian gravity or the theory of Electromagnetism has two parts that describe it:

1. **Field equation**

   A field equation is an equation of motion that describes exactly how a certain field is determined by some set of sources. In other words, it is usually a second order differential equation that has to be solved to determine the field from some collection of sources.

   Example 1.1. Newtonian gravity field equation

   $$\nabla^2 \Phi = 4\pi G \rho$$

   where ρ is the mass density. We solve this Laplace's equation to determine the field. If ρ is a $\nabla(\vec{r})$ function such that it describes a point mass, which implies that $\Phi \propto {}^{1}\!/_{\vec{r}}$.

   Example 1.2. Electromagnetic field equations. See the Maxwell's Equation's above. In equation (1.2), we have that ρ represents the charge density to determine the field in terms of the forces from some collection of charge sources.

2. **Force law**

The force law is an equation that determines how an object moves in the presence of a field. So we start by describing a collection of sources and how they describe some force field, and then we use the force law to determine how bodies in motion are affected by the force field using the force law.

Example 1.3. Newtonian force law

$$\vec{F}(t,\vec{x},\dot{\vec{x}}) = \frac{m}{2}\frac{d}{dt}\frac{\partial}{\partial\vec{v}}\vec{v}^2 = m\dot{\vec{v}} = -m\vec{\nabla}\Phi(t,\vec{x})$$

where $\dot{\vec{v}} = {}^{d^2\vec{x}}/_{dt^2} = \ddot{\vec{x}}$. This force law describes how some gravitational potential affects the motion of bodies in the presence of the field described by the potential.

Example 1.4. Lorentzian force law

$$\vec{F}\left(t,\vec{x},\dot{\vec{x}}\right) = q(\vec{E} + \vec{v}\times\vec{B})$$

where $\vec{v} = {}^{d\vec{x}}/_{dt} = \dot{\vec{x}}$.

The important thing here is that fields are just functions of points in spacetime. For example, the gravitational potential $\Phi(t,\vec{x})$ yields a number that depends on where you are in space time, and the electric field $\vec{E}(t,\vec{x})$ is also a vector that depends on where you are in space and time. We should think of force laws as equations that describe how the motion of objects deviate from being "straight lines".

To emphasize, let us assume that an object in inertial motion on the $+\hat{x}$ axis, neglecting field interactions with the object. Evidently so, it's acceleration is zero, yielding the differential equation $\ddot{x} = 0$. The solution to this second order differential equation yields motion in a straight line. So the force law considering a field with an object experiencing this motion will be able to distort the object's motion such that it is not a straight line anymore. For instance, a gravitational field will bend the straight line into a curve as it generates acceleration in the $-\hat{y}$ axis, leading to parabolic motion on the $\hat{x}\hat{y}-$plane.

Definition 1.1. General Relativity is a complete reinterpretation of gravitation such that it is not a field using a potential $\Phi(t,\vec{x})$, but instead it is a feature of spacetime itself. In particular, we replace the gravitational potential with a *metric tensor* that describes this feature, particularly, the geometrical curvature of spacetime.

We will be learning pseudo-Riemann topological spaces by employing Einstein's field equation, which determines how spacetime curvature is determined in the presence of matter or energy. In Newtonian gravitation, the source term was a mass (or energy density in Special Relativity where $E = mc^2$). Recall however that mass has no independent meaning in terms of relativity as energy and momentum both depend on the reference frame where they are measured. The "source" of the curvature in general relativity is a term in the Einstein field equations that describes a generalized mass-energy distribution in spacetime of the sources present.

The force law is also replaced by the geodesic equation, which tells us how objects move through some curved spacetime. Particularly, the geometric interpretation of the geodesic is quite simple; it is the statement that *object will move on **geodesics**, which are the shortest paths between two points on a surface.* In a flat surface, this is a straight line, but in curved surfaces, this trajectory will not be a straight line, like two points on the surface of a sphere.

(a) Geodesic over a flat surface

(b) Geodesic over a spherical surface

Figure 1: Examples of geodesic on different surfaces.

So in other words, it is the curve that minimizes the distance between two points over some surface. The force law is therefore just the statement that bodies will move in geodesics over some geometry, even when said geometry is curved. So when we think of gravity as the curvature of spacetime, then geodesics will describe motion over spacetime as it is curved.

1.2 Special Relativity

The discussion of special relativity will serve to introduce spacetime properly. We will develop it further later on, but for now we will proceed with a heuristic manner.

Definition 1.2. **Spacetime** is a smoothly connected manifold where the points defined are called **events**.

The events in spacetime can be smoothly parametrized using coordinates, such as the most common system of coordinates used, the Cartesian coordinates.

Example 1.5. Cartesian coordinates

$$(t, \vec{x}) = (-c\hat{t}, \hat{x}, \hat{y}, \hat{z})$$

In relativistic mechanics, we usually consider all directions of motion in a single vector $\vec{x} = (\hat{x}, \hat{y}, \hat{z})$ while time is multiplied by the negative speed of light $t = (-c\hat{t})$. We will explain this later on.

Thus, spacetime is a smooth manifold with events as points parametrized by some coordinates, and in relativity, we know that physics should be independent of its coordinate system. This is known as the **principle of general covariance**.

To begin, let us assume we are working with Newtonian mechanics where we have two points in spacetime labelled $(t_1, \vec{x}_1) = (t_1, x_1, y_1, z_1)$ and $(t_2, \vec{x}_2) = (t_2, x_2, y_2, z_2)$ respectively, both of which describe a single event in spacetime given as

$$(\Delta t, \Delta \vec{x})$$

In this event, Δt denotes the **temporal separation** while $\Delta \vec{x}$ denotes the **spatial separation**. Explicitly:

$$\begin{aligned}
\Delta t &= t_2 - t_1 \\
\Delta \vec{x} &= \sqrt{(\vec{x}_2 - \vec{x}_1) \cdot (\vec{x}_2 - \vec{x}_1)} \\
&= \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2} \qquad \text{(in Cartesian coordinates)}
\end{aligned}$$

This makes sense as Newtonian mechanics can always be written in terms of both Δt and $\Delta \vec{x}$, so the numerical value of t_1 will depend on the origin in the frame of reference we are working on. In other words, if we choose our reference frame to be Montreal at 6:00pm and remain at rest until 6:01pm, then we can denote that $\Delta t = 60$ s. So if we now jog in Montreal at constant velocity of 1 m/s in this reference frame, the velocity will depend on the time separation Δt and not on the absolute values of t_1 and t_2. So, if we suddenly change the time frame to be 4:20pm to 4:21pm, the velocity should not vary due to the absolute value of the new points in time, as the time separation Δt would be the equal to the case at 6:00pm. The same would go for the choice of spatial coordinates, so if we move from Montreal to Toronto and the speed remains constant, the space separation $\Delta \vec{x}$ should also not vary. To summarize, the laws of physics would remain invariant even if we change the choice of reference frame where the choice of time is independent of the choice of space, and so they are independent of the entire coordinate system chosen. Albeit elementary to classical mechanics, it is vital to understand this properly to proceed with the philosophy of relativity.

In special relativity, there is no separate notion between space and time, so we denote both as spacetime. This means that the both the separation of time and the separation of space are dependent on each other, which leads to effects like time dilation. Time dilation implies that *the temporal difference Δt of two event in spacetime will depend on the relative position in spacetime where an observer is viewing the event from.* Similarly, the distance or length of an object will depend on where the observer is in spacetime. In other words, we cannot think of Δt and $\Delta \vec{x}$ as independent separations in spacetime, as one is directly related to the other depending on the frame of reference. So, observing the same event in a different reference frame will change the values of Δt and $\Delta \vec{x}$.

However, we can denote an invariant notion to both the spatial and temporal separations called the **interval** between two events as

$$\Delta s^2 = -c^2 \Delta t^2 + \Delta \vec{x}^2 \qquad \text{(in Cartesian coordinates)}$$

Usually, we simplify by allowing $c = 1$ in units in terms of the speed of light [3×10^8 m/s] such that

$$\Delta s^2 = -\Delta t^2 + \Delta \vec{x}^2 \tag{1.6}$$

So when we measure distance with respect to the speed of light, we do so in terms of light-time, such as light-seconds and light-years. It is usually helpful to draw pictures in a spacetime diagram.

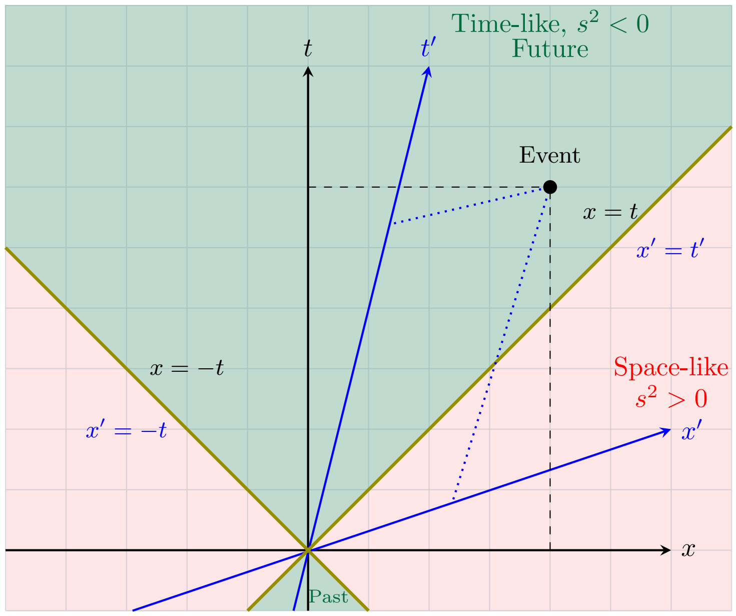

Figure 2: A spacetime diagram of an time-like event with the light cone in two different reference frames, a static frame $K = (t, x)$ and another moving away from the static frame at a fraction of the speed of light, $K' = (t', x')$. Notice how even if the spatial ($\Delta x \neq \Delta x'$) and temporal ($\Delta t \neq \Delta t'$) separations vary from one reference frame to the other, the interval in spacetime remains invariant $\Delta s^2 = \Delta(s')^2$. We usually denote $(t_1, x_1) = (t_1', x_1') = (0, 0)$ for convenience in either reference frame.

Notice in the spacetime diagram that when $\Delta s^2 = 0$, we have that $\Delta x = \pm \Delta t$, defining a **light cone**. Events defined on the light cone are denoted **light-separated events** or **null-separated events**. As otherwise noted, events where $\Delta s^2 < 0$ are inside the light cone and denoted as **time-like** or **time-separated**, while events where $\Delta s^2 > 0$ lie outside the light cone and are denoted as **space-like** or **space-separated**. Notice that for time-like events, it is clear that when $t_2 > 0$, the event will propagate to the future, while if $t_2 < 0$ it propagates to the past. Also note that space-like events are deemed as impossible, as there is no geodesic inside a light cone that can connect the origin to the event. In other words, an object would have to travel faster than light to arrive at this location in spacetime, which is indeed impossible in special relativity.

We would get hyperbolic equilibrium to a 2D spacetime system with respect to the interval Δs^2. This means a time-like event will always be either time-like or get close to null-like, while a space-like event will also either remain space-like or get close to null-like.

Claim. *We can reduce Special Relativity to the statement that for 2 time-like separated events, the proper time $\Delta\tau$ measured by an observer moving at constant velocity between the two events is given by:*

$$(\Delta\tau)^2 = -(\Delta s)^2$$

Example 1.6. Time dilation

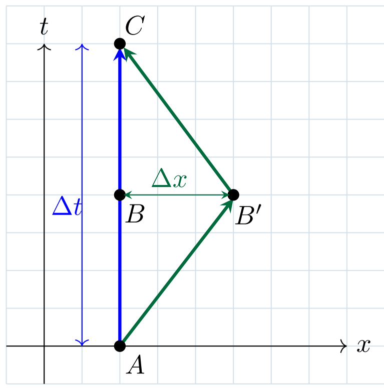

Figure 3: Spacetime diagram for the twin's paradox. Notice that the inclination of the trajectory of the second twin is greater than $\pi/4$ with respect to the x-axis, implying that the speed $v < c$.

Here, we see the famous *twin's paradox.* We have two observers relative to a frame of reference, usually considered twins, that will go from a point $A(t, x) = (0, 0)$ to a point $C(t, x) = (\Delta t, 0)$ in spacetime. The first twin will remain static in space while the second will go to point $B'(t, x) = \left(\frac{\Delta t}{2}, \Delta x\right)$ before reaching point C, where $\Delta x = \frac{1}{2}v\Delta t$ for constant propagation velocity v. The proper time for the trajectory of the first twin from A to B is clearly $\tau_{B,A} = \frac{\Delta t}{2}$, but for the second twin, the proper time from A to B' is given by

$$\tau^2_{B',A} = \left(\frac{\Delta t}{2}\right)^2 - (\Delta x)^2$$

$$\therefore \boxed{\tau_{B',A} = \frac{\Delta t}{2}\sqrt{1 - v^2}} \tag{1.7}$$

By symmetry, it follows that $\tau_{C,B} = \tau_{B,A}$ and $\tau_{B',A} = \tau_{C,B'}$. Considering the entire sequence of movement for both twins, we have

$$\tau_{C,B,A} = \Delta t$$
$$\tau_{C,B',A} = \sqrt{1 - v^2}\Delta t$$

Clearly, $\tau_{C,B,A} > \tau_{C,B',A}$, and so the second twin is younger than the first twin by the time they both reach the point C. This implies that the proper time τ along some world line is related to the interval Δs^2 traversed through spacetime. Therefore, we can assess the ratio of relative time difference between the twins as

$$\frac{2\tau_{C,B',A}}{\tau_{C,B,A}} = \sqrt{1 - v^2}$$

Example 1.7. Length contraction

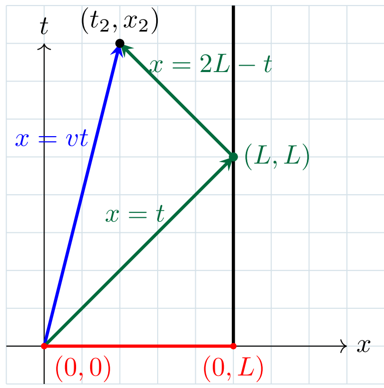

Figure 4: The spacetime diagram for length contraction.

Consider an observer moving at constant velocity v from the origin of a reference frame with respect to a static ruler with a light bulb at the point $(0, 0)$. We would like for the observer to measure the ruler using light-travel time, which is achieved by flashing the light bulb and measuring the time it takes to reach a point (L, L), and return to the observer at the intersection point (t_2, x_2). Noting how we have two solutions for x, namely $x = 2L - t$ and $x = vt$, we can equate them to find (t_2, x_2) such that

$$t_2 = \frac{2L}{(1+v)}$$
$$x_2 = v\left(\frac{2L}{(1+v)}\right)$$

And so, the proper time $\Delta\tau^2$ measured by the observer at the point when the light bulb's ray is reflected back is given by $\Delta\tau^2 = \Delta t^2 - \Delta x^2$ such that

$$\Delta\tau^2 = \left(\frac{2L}{(1+v)}\right)^2 (1 - v^2)$$

Knowing that an observer at rest with the ruler Δt^2 for the light to reflect back on itself is $\Delta t = \frac{2L}{(1+v)}$, we get the ratio

$$\frac{\Delta\tau^2}{\Delta t^2} = (1 - v^2)$$

So the apparent length of the ruler would be shorter by a factor of v^2. Note that in the case where $x = -vt$ for the observer's trajectory, the same ratio would arise.

We are very habituated to writing the interval in the following way

$$\Delta s^2 = -\Delta t^2 + \Delta x_1^2 + \Delta x_2^2 + \Delta x_3^2$$

We will consider the x^μ notation, where $\mu = 0, 1, 2, 3$, to denote these coordinates as a single coordinate. Moreover, we can denote x^i where $i = 1, 2, 3$ as the spatial coordinates. Therefore, we may define

- $x^0 = t$
- $x^2 = x_2$
- $x^1 = x_1$
- $x^3 = x_3$

We can thus summarize the coordinates of some interval to a summation

$$\Delta s^2 = \sum_{\mu=\nu} \eta_{\nu\mu} \Delta x^\mu \Delta x^\nu$$

Where we define $\eta_{\mu\nu}$ to be a 4×4 diagonal matrix such that

$$\eta_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Definition 1.3. The **metric** of flat spacetime in Cartesian coordinates is given by the 4×4 matrix $\eta_{\mu\nu}$.

We may further simplify the interval by using *Einstein summation notation.* This notation takes the summation symbol for granted, which adds all terms in the resulting 4×4 matrix whose indices follow the property $\mu = \nu$, to get

$$\Delta s^2 = \eta_{\mu\nu} \Delta x^\mu \Delta x^\nu \tag{1.8}$$

As such, in tensor calculus language this implies that for every subscript in a tensor we must have a corresponding superscript to yield an interval as desired. In this case, $\eta_{\mu\nu}$ has two subscripts, each denoting an index for the matrix, and the one-forms Δx^μ and Δx^ν each have one superscript to indicate their contravariant components. Hence, we can summarize special relativity to the statement that all intervals in spacetime can be given by an equation in the form of (1.8), such that spacetime is denoted as **Minkowski spacetime** $\mathbb{R}^{1,3}$ or simply $\mathcal{M}$. To summarize for now, general relativity implies that spacetime can be something other than Minkowski spacetime $\mathcal{M}$.

Other examples of spaces are Euclidean spaces, both in 2D ($\mathbb{R}^2$) and in 3D ($\mathbb{R}^3$), where an invariant interval is simply denoted as the distance between two points. Using the notation we have established, the coordinates in 2D and 3D respectively would be:

$$(x^1, x^2) = (x, y)$$
$$(x^1, x^2, x^3) = (x, y, z)$$

For either of these, the distance is obtained by use of Pythagoras's theorem $D = \sqrt{(x^1)^2 + (x^2)^2 + (x^3)^3}$, where in 2D we simply neglect the x^3 term. Recall than an invariant interval is similarly given by $\Delta s^2 = -(x^0)^2 + (x^1)^2 + (x^2)^2 + (x^3)^3$. where the x^0 term is given a negative sign. Note that an opposite yet equally valid for the invariant interval is $\Delta s^2 = \Delta t^2 - \Delta x^2$. This convention is better for events happening in the same spatial coordinate at different points in time, while the one we used is better for events that happen at the same time in different spatial coordinates.

Exercise 1.1. Recalling that a Lorentzian transformation Λ is a linear change of coordinates, determine the correct form of the 2×2 matrix Λ that validates the following change of coordinates:

$$\left[\begin{smallmatrix} x \\ t \end{smallmatrix}\right] \to \left[\begin{smallmatrix} x' \\ t' \end{smallmatrix}\right] = \Lambda \left[\begin{smallmatrix} x \\ t \end{smallmatrix}\right]$$

To summarize, the proper time $\Delta\tau$ measured by an observer moving at some constant velocity v between two points in spacetime is given by

$$(\Delta\tau)^2 = -(\Delta s)^2 = \Delta t^2 - \Delta\vec{x}^2$$

Where we could write the invariant interval Δs^2 using Einstein summation notation as

$$\Delta s^2 = \eta_{\mu\nu}\Delta x^\mu \Delta x^\nu$$

For the coordinates $\Delta x^\mu = (\Delta t, \Delta\vec{x})$ where $\mu = 0, 1, 2, 3$, and the metric $\eta_{\mu\nu}$. Recall that this notation neglects a summation symbol that will sum all elements in the 4×4 matrix $\eta_{\mu\nu}\Delta x^\mu \Delta x^\nu$ that follow the property $\mu = \nu$.

Thus far, we have only considered an observer moving at constant velocity for the sake of conceptual simplicity. However, when consider gravitation we will encounter accelerating reference frames. That is, we will encounter motion in spacetime that is not given by a straight line, but rather by some curve, implying variable velocity.

Definition 1.4. A general path through spacetime is denoted as a **worldline** and can be parametrized with 4 functions, one for each coordinate in the metric.

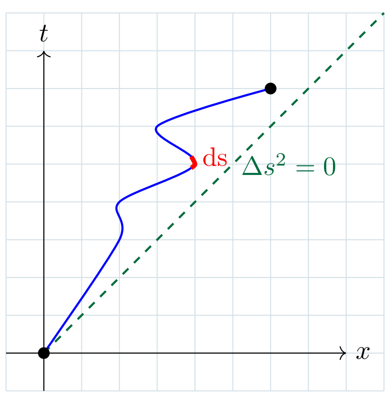

Figure 5: A worldline in spacetime of a time-like observer accelerating, with a line element shown. Notice how at every point on the worldline, the observer's speed has the property $v < c$.

The parametrization in Cartesian coordinates for a worldline using λ as the variable which denotes points along the world line is as follows:

$$x^\mu(\lambda) = \big(t(\lambda), x^1(\lambda), x^2(\lambda), x^2(\lambda)\big)$$

Along this worldline, we consider infinitesimally small displacements through space $d\vec{x}$ and through time dt such that at each interval of displacement we approximate to constant velocity. The proper time $\Delta\tau$ measured by such an accelerating observer along a worldline is hence given by $\sqrt{-(\Delta s)^2}$, where

$$\Delta s = \int \sqrt{-\left(\frac{dx^0}{d\lambda}\right)^2 + \left(\frac{dx^1}{d\lambda}\right)^2 + \left(\frac{dx^2}{d\lambda}\right)^2 + \left(\frac{dx^3}{d\lambda}\right)^2}\, d\lambda$$

Which is simply the arc length of a curve, as expected. We could also write this using Einstein summation convention to make it cleaner:

$$\Delta s = \int \sqrt{\eta_{\mu\nu}\frac{dx^\mu}{d\lambda}\frac{dx^\nu}{d\lambda}}\, d\lambda$$

For convenience, we will denote either of the above as

$$\Delta s = \int ds$$

Where we define ds as the **line element** of a worldline, given by

$$\boxed{ds^2 = \eta_{\mu\nu} dx^\mu dx^\nu} \tag{1.9}$$

We think of ds as an infinitesimal variation on the interval Δs, and so dx^μ and dx^ν are infinitesimal displacements on the coordinates x^μ.

We again redefine Special Relativity to the statement that the line element is given by equation (1.9), where we defined $\eta_{\mu\nu}$ as a metric given by a 4×4 diagonal matrix with constant components. In General Relativity we have the statement that all of the effects of gravity are packaged in terms of a replacement

$$\eta_{\mu\nu} \to g_{\mu\nu}(x)$$

Which is a 4×4 matrix, formally called a tensor, whose components can change according to the value of the coordinates given, rather than remaining constant as in Special Relativity. we denote $g_{\mu\nu}(x)$ as the **metric of curved spacetime**. This metric is parametrized by 16 functions of the coordinates of spacetime; one parametrization per component in the matrix. So the line element in General Relativity changes to

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu$$

Example 1.8. Light-rays in General Relativity

When we consider General Relativity, light propagation still takes the form of a *null-like* worldline. However, the definition of null-like worldline, considering $ds^2 = 0$.

Example 1.9. Massive objects in General Relativity

Massive objects in general relativity travel along *time-like* worldlines such that $ds^2 < 0$, implying that their speed along a world line has the property $v < c$ at every point. This implies that the proper time $\Delta\tau^2$ measured by an observer in GR travelling along some worldline $x^\mu(\lambda)$ is given by

$$\Delta\tau^2 = -\Delta s^2$$

Where we have

$$\Delta s = \int ds = \int \sqrt{g_{\mu\nu} \frac{dx^\mu}{d\lambda} \frac{dx^\nu}{d\lambda}} d\lambda$$

We can now see why we say that Special Relativity is just the case of General Relativity where we will have a constant metric. And so, every case in Special Relativity can be solved with this form of a line element. Recall that all of this merely serves as a review of the study of Special Relativity such that we may introduce General Relativity, so we still need to formalize what has been stated.

1.3 Equivalence Principle

We will now introduce the **equivalence principle**. Recall the motion of objects in the presence of a gravitational potential in Newtonian gravitation:

$$\vec{F}(\vec{v}) = m_i \dot{\vec{v}} = -m_g \vec{\nabla} \Phi$$

We usually assumed that the inertial mass m_i, the mass due to the forces interacting with an object, and the gravitational mass m_g, the mass due to the gravitational potential interacting with the object were equivalent, and approximately speaking in human scales, they are. However, it is not true that we can say this generally. For every object and type of matter we know, both m_i and m_g are equal, but we do not know if all unknown objects will follow this rule. Given current evidence, we can cancel both masses out to assume that, for all objects known,

$$\dot{\vec{v}} = \vec{a} = -\vec{\nabla} \Phi$$

So the acceleration of an object will be independent of its mass, and as stated before, this is considered as an observational fact. In a gravitational potential, there are trajectories defined by the solutions to the equations of motion due to the potential on some object, denoted as **free-falling** or **inertial** trajectories, where no forces are interacting with an object. In a free-falling object, the object only experiences gravitational potential.

Now, consider an infinitesimally small region where $\vec{\nabla}\Phi \approx -\vec{a}_0$, where $\vec{a}_0$ is some constant to within some approximation made by measurements.

Definition 1.5. The **weak equivalence principle** states that in this small region, the effects of gravity are indistinguishable from the effects of a constantly accelerating reference frame.

Example 1.10. Acceleration Experiment

Let us imagine that we are in a laboratory on Earth as our frame of reference, which is very small. We can thus approximate the gravity with a downwards pull in this laboratory to be constant to some significant figure. The downwards gravitational force we experience within the lab is given by

$$\vec{F}_g = -mg = -G\frac{mM_E}{R_E^2}$$

Where $G \approx 6.7 \times 10^{-11}$ ${}^{\mathrm{N \cdot m^2}}/_{\mathrm{kg^2}}$ is the gravitational constant, $M_E \approx 6.0 \times 10^{24}$ kg is the mass of the Earth and $R_E \approx 6.4 \times 10^{7}$ m is the radius of the Earth as measured in this lab, so we assume that $g = 9.8$ ${}^{\mathrm{m}}/_{\mathrm{s^2}}$.

We can write this in a more mathematical manner as

$$\ddot{x} = -g = -9.8 \ ^{\mathrm{m}}\!/\!_{\mathrm{s}^2}$$

Now, assume that we take this laboratory into a spaceship such that the laboratory is experiencing no gravitational force from Earth. However, the spaceship that contains the lab is experiencing upwards acceleration of $a_0 = g = 9.8 \ ^{\mathrm{m}}\!/\!_{\mathrm{s}^2}$, and thus so is the laboratory. This would imply that the normal acceleration experienced by us inside the lab would be $a_n = -g = -9.8 \ ^{\mathrm{m}}\!/\!_{\mathrm{s}^2}$. Clearly, this would mean that the effects of gravity would be indistinguishable from those of the accelerating spaceship if we were to perform mechanical experiments. In fact, pilots use this technique to accelerate upwards and downwards to simulate zero gravity. You can watch Stephen Hawking in such a plane ride here[1]. In more mathematical sense, we notice that the spaceship frame of reference is not inertial. The coordinates we define are thus accelerating when defined. To find the new x', we first consider constant acceleration:

$$\dot{x}\frac{dx}{dt} = -g$$

$$\therefore \dot{x}\int_0^1 dx = -g\int_0^t dt$$

We are considering the initial position in spacetime x and wish to find a new position x' relative to the acceleration. We integrate to find their difference with respect to time such that

$$\int_x^{x'} dx = -g\int_0^t t dt$$

$$\therefore x' = x - \frac{1}{2}gt^2$$

Where we have that (t, x) are the usual spacetime coordinates outside the the laboratory and (t, x') are the coordinates inside the lab as it is accelerating upwards. Hence, it becomes clear that

$$\ddot{x} = \ddot{x}' = -g$$

Definition 1.6. The forces generated due to acceleration are denoted as **fictitious forces**.

The main idea behind the equivalence principle is that inside a small frame of reference, the forces of Earth's gravity are roughly equivalent to the fictitious forces of the same frame accelerating in spacetime at $9.8 \ ^{\mathrm{m}}\!/\!_{\mathrm{s}^2}$. We can extend this to other gravitational sources aside from the Earth. As a result, we have that the metric of Minkowski space $\eta_{\mu\nu}$ will no longer be constant in terms of the accelerating frame of reference, and moreover, it implies that gravity itself is a result of the variance of the metric.

[1] Please forgive the potato quality as it is a 14 year old video as of the time of this writing.

Once we assume that gravitational potential will vary over a larger region than we defined on the previous reference frames, for example in an entire planet or a star, we cannot think of gravity as a fictitious force. So imagine that the mighty nation of Canada decided to build a colossal completely flat spaceship of around 200 km in length over the continental United States, such that the corners of the ship are above the surface by around 800 m, while the centre rests in contact with the surface. Thus, if we measure the gravitational force of the Earth along various points along the spaceship, we will find that the gravitational force is not constant if we are being precise. If we then reach space and accelerate the entire ship upwards relative to the Earth's frame of reference at exactly 9.8 $^{\mathrm{m}}/_{\mathrm{s}^2}$, we cannot say that the accelerating frame of the ship has an equivalent fictitious force to that of the gravity of Earth, as there are several points where gravity will not be equal.

Definition 1.7. The **strong equivalence principle** or **Einstein equivalence principle** states that in a small enough frame of reference, the laws of physics reduce to those of special relativity. Moreover, spacetime locally looks like Minkowski space.

Once we review differential geometry and tensors, we will be able to redefine this principle yet again to a much more mathematical version:

For any point given by some coordinates x_0^μ in space time, there exists a coordinate system such that the line element takes that of special relativity near that specific point x_0^μ.

$$ds^2 = \eta_{\mu\nu} dx^\mu dx^\nu + \mathcal{O}(x - x_0^\mu)$$

Where $\mathcal{O}(x - x_0^\mu)$ denotes the error in regard to the extent at which gravity is not a fictitious force.

Exercise 1.2. Determine the expression for a line element ds^2 in terms of the accelerating coordinates in Example 1.10, as given by

$$x' = x - \frac{1}{2}gt^2$$
$$t' = t$$

2 Manifolds

We recall the equivalence principle which states that spacetime locally looks like Minkowski space $\mathbb{R}^{3,1}$. In small frames of reference we can presume that gravity can be ignored and thus assume a pure Minkowski space with accelerating frames of reference. In larger frames of reference, gravity does not act as a fictitious force, but the knowledge that it indeed is equivalent to a fictitious force is important nonetheless. To understand this. take the coordinate system we usually employ on the Earth itself of latitude and longitude. When we consider a local area as seen by a human being, these coordinates can be presumed to be in a flat space. However, when considering the same coordinates over larger distances as when travelling in a Trans-Atlantic flight, the metric describing the coordinates curves with the Earth itself.

Theorem 2.1. *Spacetime is a 4-dimensional pseudo-Riemann manifold.*

Because we are physicists, we will not prove this statement and instead consider it to be true while justifying it along the way with various concepts. For a more thorough analysis a book on differential geometry is highly recommended, but is not necessary for the following content.

Definition 2.1. A D-dimensional **manifold** M is a set of points that can be labelled by coordinate systems in a smooth way.

We will assume all manifolds in this context to be smooth henceforth.

Definition 2.2. A **coordinate system** is a set of D functions x^i which label different points along the manifold M in some region $U \subset M$ in a unique way. More precisely,

$$x^i : U \to \mathbb{R}^D \quad \text{for } i = 1, \ldots, D$$

Definition 2.3. The region U which the coordinates label uniquely is called **coordinate patch** of some manifold M. We say that the manifold M is covered by the coordinates x^i.

Definition 2.4. Global coordinate systems are the coordinate systems which cover all of M such that $U \equiv M$.

Generally, many coordinate systems cover some patch $U \subset M$ while not all of M and remain useful in various cases.

Example 2.1. Assume that we have two coordinate systems x^i and x^j where $i, j = 1, \ldots, D$ which correspond to different ways of labeling a patch $U \subset M$. We can then write the coordinate transformation between x^i and x^j in two ways:

$$x^i(x^j) \quad \text{or} \quad x^j(x^i)$$

Where both of these are infinitely differentiable functions everywhere on the manifold M.

The first observation we make is the following:

Theorem 2.2. *If some two coordinate systems x^i and x^j uniquely label some points in a patch $U \subset M$, then we know that the Jacobian matrix J of this coordinate transformation is invertible, where*

$$J = \frac{\partial x^i}{\partial x^j}$$

Corollary 2.3. *The determinant of an invertible matrix is nonzero. Therefore,*

$$\det(J) \neq 0$$

Corollary 2.4. *If $\det(J) = 0$ then one of the coordinate systems is implied to be* ***degenerate****, or* ***singular****, such that it no longer labels points in $U \subset M$ uniquely.*

Example 2.2. 2D Manifold

Consider the manifold $M = \mathbb{R}^2$. We define two different sets of coordinates given by

$$\begin{aligned} (x, y) &\quad \text{Cartesian coordinates} \\ (r, \theta) &\quad \text{Polar coordinates} \end{aligned}$$

We find that the Cartesian coordinates are global for the manifold M, as any point in the manifold can be described by a point in Cartesian coordinates. However, the Polar coordinates are not global as they label all points in the Manifold except the origin. Notice that at the Cartesian origin $x = y = 0$ it follows in the change of coordinates that θ no longer parametrizes the manifold, and so the coordinate system degenerates into r only. To see this, let us set up the coordinate systems such that

$$\begin{aligned} x^i &= (x^1, x^2) = (x, y) \\ x^j &= (x^{1'}, x^{2'}) = (r, \theta) \end{aligned}$$

We can therefore write $x^i(x^j)$ and $x^j(x^i)$as

$$\begin{aligned} x^1(x^j) &= x^{1'} \cos x^{2'} = r\cos\theta & \qquad x^{1'}(x^i) &= \sqrt{(x^1)^2 + (x^2)^2} = \sqrt{x^2 + y^2} \\ x^2(x^j) &= x^{1'} \sin x^{2'} = r\sin\theta & \qquad x^{2'}(x^i) &= \arctan\left(\frac{x^2}{x^1}\right) = \arctan\left(\frac{y}{x}\right) \end{aligned}$$

Now, we build the Jacobian $J = \frac{\partial x^i}{\partial x^j}$ and take its determinant:

$$\det(J) = \begin{vmatrix} \frac{\partial x^1}{\partial x^{1'}} & \frac{\partial x^1}{\partial x^{2'}} \\ \frac{\partial x^2}{\partial x^{1'}} & \frac{\partial x^2}{\partial x^{2'}} \end{vmatrix} = \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{vmatrix} = \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix}$$

$$\therefore \det(J) = r^2\left(\cos^2\theta + \sin^2\theta\right) = r^2$$

We see that the θ components get cancelled such that $r^2 = 0 \implies \det(J) = 0$, as if the determinant is zero we have that $\cos^2\theta + \sin^2\theta \neq 0$. This effectively shows that Polar coordinates are not global coordinates, as they degenerate at the origin. However, we still find a lot of use for this coordinate system.

Exercise 2.1. Compute the Jacobian $J' = \frac{\partial x^j}{\partial x^i}$ and find its determinant. Discuss the difference between both Jacobians J and J'.

When describing spacetime, we use a special notation for the coordinates in general relativity. We write x^μ for $\mu = 0, 1, 2, 3$ when considering a spacetime manifold rather than x^i for $i = 1, 2, 3$ when considering Euclidean manifolds.

Example 2.3. cartesian coordinates of flat Minkowski spacetime

$$x^\mu = (t, \vec{x}) \in \mathbb{R}^{1,3}$$

2.1 Principle of General Covariance

Definition 2.5. The **Principal of General Covariance** requires that we write the laws of physics in such a way that any physical prediction made by these laws is independent of the choice of coordinate system to write that law.

The statement above is incredibly important. In order to write any physical law, we need to describe a set of coordinates to denote points in spacetime. However, what we want is that the outcome of every experiment be independent of any choice of coordinates chosen. It might seem obviously intuitive, but it is rather important to understand the mathematics behind General Relativity, as Einstein's field equations arise from this very principle, and thus, so does the theory of Generalized Relativity.

The effect of the principle of General Covariance is that we need to understand now how various objects can transform under these changes of coordinates, so the laws we write must faithfully obey this principle.

Example 2.4. Line element of Minkowski space (revisited)
Recall the line element of Minkowski space:

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

Now, consider the change of coordinates $x^\mu \to x^\alpha(x^\mu)$ such that we have

$$dx^\mu = \left(\frac{\partial x^\mu}{\partial x^\beta}\right) dx^\beta$$

Recall that as we are using Einstein notation, the differential of the first coordinate system is indeed a set of

4 equations where $\mu = 0, 1, 2, 3$ such that

$$\boxed{dx^\mu = \sum_{\beta=0}^{3} \left(\frac{\partial x^\mu}{\partial x^\beta}\right) dx^\beta = \left(\frac{\partial x^\mu}{\partial x^\beta}\right) dx^\beta} \tag{2.1}$$

Which in total implies 16 terms, 4 per equation. Recall the Cartesian and Polar coordinates in a 2D manifold given by

$$x = r\cos\theta$$
$$y = r\sin\theta$$

The differentials of the Cartesian coordinates as given by equation (2.1) in this case are

$$dx = \left(\frac{\partial r\cos\theta}{\partial r}\right) + \left(\frac{\partial r\cos\theta}{\partial\theta}\right) = \cos\theta dr - r\sin\theta d\theta$$
$$dy = \left(\frac{\partial r\sin\theta}{\partial r}\right) + \left(\frac{\partial r\sin\theta}{\partial\theta}\right) = \sin\theta dr + r\cos\theta d\theta$$

Without loss of generality, the line element of flat Minkowski space in terms of some coordinates x^α to change from some given coordinates x^μ would be given as

$$\begin{aligned} ds^2 &= \eta_{\mu\nu} dx^\mu dx^\nu \\ &= \left(\frac{\partial x^\mu}{\partial x^\alpha} dx^\alpha\right)\left(\frac{\partial x^\nu}{\partial x^\beta} dx^\beta\right)\eta_{\mu\nu} \end{aligned}$$

We now rewrite the coordinate change with the metric $\eta_{\mu\nu}$ itself to get

$$\boxed{g_{\alpha\beta} = \eta_{\mu\nu}\frac{\partial x^\mu}{\partial x^\alpha}\frac{\partial x^\nu}{\partial x^\beta}} \tag{2.2}$$

Which yields the cleaner version of the line element as denoted by

$$\boxed{ds^2 = g_{\alpha\beta} dx^\alpha dx^\beta = \eta_{\mu\nu} dx^\mu dx^\nu} \tag{2.3}$$

we are hence given the metric transformation $\eta_{\mu\nu} \to g_{\alpha\beta}$. Again, recall the case of a 2D manifold. We will have the Cartesian differential

$$ds^2 = dx^2 + dy^2$$

To get the Polar differential, we apply again equation (2.1)

$$\begin{aligned} ds^2 &= (\cos\theta dr - r\sin\theta d\theta)^2 + (\sin\theta dr + r\cos\theta d\theta)^2 \\ &= dr^2 + r^2 d\theta^2 \end{aligned}$$

Exercise 2.2. Consider the analogue of Polar coordinates in Minkowski space called Rindler coordinates. Recalling the usual Cartesian coordinates in Minkowski space, we write them in terms of Rindler coordinates as follows:

$$t = r\sinh\tau$$

$$x = r\cosh\tau$$

Determine the Jacobian and find the line element in these coordinates.

2.2 Tensors

Definition 2.6. A **tensor** $T^{\nu_1,\nu_2,\cdots}_{\mu_1,\mu_2\cdots}$ is an algebraic structure that generalizes quantities by use of a multi-dimensional array. The subscripts μ_i for $i \in \mathbb{N}$ represent **covariant indices** such that each subscript is an index for covariant elements. Recall that covariant elements undergo the same transformation as the basis in a change of basis, such as basis vectors. The superscripts ν_i for $i \in \mathbb{N}$ represent **contravariant indices** such that their transformation matrix in a change of coordinates is the inverse of the change of basis matrix.

Definition 2.7. We denote types of tensors by giving them a **weight** representing the number of n directional and m covariant indices in a 2-tuple given by (n, m). We determine the **rank** r of a tensor by adding the components of the weight such that $r = n + m$.

It is important to understand how quantifiable values can transform under coordinate transformations. Consider an arbitrary D-dimensional manifold M where a patch $U \subset M$ in the manifold can be described using a multitude of coordinate systems, namely x^μ and x^α where $\mu, \alpha = 0, 1, \ldots, (D-1)$.

Definition 2.8. A **scalar function** a, or simply **scalar**, is a (0,0)-tensor a that maps elements from a field to points in a manifold M independently from the choice of coordinates.

$$a = a(x^\mu) \quad \forall x^\mu \in M$$

To change a scalar quantity from x^μ coordinates to x^α coordinates, we have

$$a^* = a\left(x^\alpha\right) = a\left(x^\alpha(x^\mu)\right) \quad \forall x^\alpha(x^\mu) \in M$$

Theorem 2.5. *A scalar function a is defined to be invariant under an invertible coordinate transformation given by $x^\mu \to x^\alpha = x^\alpha(x^\mu)$.*

Theorem 2.5 means that scalar values defined in a Manifold will not change given a change of coordinates, as they are independent of coordinates.

Definition 2.9. A **vector** $\vec{v}$ is a (1,0)-tensor on a D-dimensional manifold that is defined as D functions called **tangent vectors** or **contravariant vectors** v^μ on a **basis** $\frac{\partial}{\partial x^\mu} = \partial_\mu$ given coordinates x^μ such that $\vec{v} = v^\mu \partial_\mu$. It transforms under the change of coordinates $v^\mu \to v^{\mu'}(x^{\mu'})$ as given by

$$v^{\mu'}(x^{\mu'}) = \frac{\partial x^{\mu'}}{\partial x^\mu} v^\mu \left(x^{\mu'}(x^\mu)\right)$$

Theorem 2.6. *The transformation on a vector v^μ is defined to be covariant under an invertible coordinate transformation given by*

$$v^\mu(x^\mu) \to v^{\mu'}\left(x^{\mu'}(x^\mu)\right)$$

Theorem 2.6 means that we will be able to represent the same vector in different coordinate systems. Thus, a change in coordinates might change its components, but the vector as an object will remain as is. It is also important to note that a vector as defined here is a generalization of the algebraic vectors presented in linear algebra.

Example 2.5. Rotations in 3D

Consider a vector $\vec{v} \in \mathbb{R}^3$ represented in Cartesian coordinates such that

$$\vec{v} = v^1\partial_1 + v^2\partial_2 + v^3\partial_3$$

We now wish to rotate the coordinate system using an orthogonal matrix[2] R such that we get a new coordinate system:

$$\vec{v} = \begin{pmatrix} x^{1'} \\ x^{2'} \\ x^{3'} \end{pmatrix} = R_{3\times 3} \cdot \begin{pmatrix} x^1 \\ x^2 \\ x^3 \end{pmatrix}$$

Using the Einstein summation notation, we let $\vec{v} = v^i \partial_{x^i}$ be the vector under the original coordinates and let $\vec{v} = v^{i'}\partial_{i'} = R^{i'}_j v^j \partial_j$ be the vector considering the new rotated coordinates. Including the summation, this is evidently equivalent to doing matrix multiplication:

$$\vec{v} = \sum_{i'=1}^{3} v^{i'} \frac{\partial}{\partial x^{i'}} = \sum_{i',j=1}^{3} \left(R^{i'}_j \cdot v^j \frac{\partial}{\partial x^j}\right)$$

[2]Orthogonal matrices A are square matrices defined with the property $A^T = A^{-1}$.

Notice how in this case, the rotation matrix indeed is the Jacobian matrix for the coordinate change as denoted in equation (2.1) such that

$$R^{i'}_{j} = \frac{\partial x^{i'}}{\partial x^{j}}$$

So the change of coordinates in a rotation is a linear transformation where the Jacobian is a matrix of constants. Furthermore, when we have a transformation of coordinates $x^j \to x^{i'}$, the indices of the transformation's domain are considered as covariant indices while the indices in the transformation's codomain remain contravariant. This is what yields i' as a contravariant index and j as a covariant index in the Rotation matrix $R^{i'}_{j}$.

Exercise 2.3. The rotation presented above is noted as a *passive rotation*. This means that the vector will not rotate on the manifold, but its coordinates will. The converse would be an *active rotation* where the vector itself is rotated yet its coordinates remain intact. Show how an active rotation on the same manifold $M = \mathbb{R}^3$ would look like and prove that the coordinates would remain unchanged.

The introduction of linear transformations into tensors will allow us to extend transformations beyond linearity to include non-linear transformations as well. Recalling that the weak equivalence principle states that in small frames of reference gravity is equivalent to a fictitious force. In other words, gravity is indistinguishable from an accelerating frame of reference, if the frame of reference is small enough.

Definition 2.10. A **non-inertial frame of reference** is a reference frame whose velocity varies over time.

Theorem 2.7. *For a non-inertial frame of reference K', the accelerating coordinates $x^{\mu'}$ of the frame are non-linearly related to the coordinates x^{μ} of an observer's reference frame K. We note that K can be inertial or non-inertial.*

Even though the components of a certain vector $\vec{v}$ in a certain coordinate system x^{μ} will look different in another coordinate system $x^{\mu'}$, the vector should be seen as an algebraic structure that is independent of the coordinates chosen. It is important to clarify that a vector has no covariant or contravariant indices, and is instead a geometric object that is present in a manifold M regardless of the coordinates chosen.

"Once you stop learning, you start dying." - Albert Einstein

Example 2.6. Worldline on a Manifold

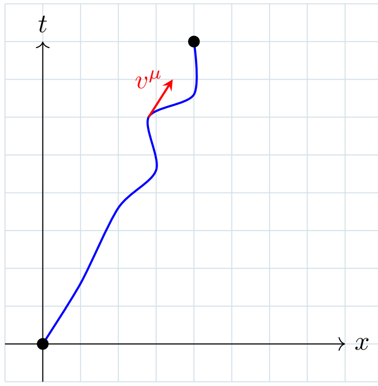

Figure 6: A worldline on a spacetime manifold parametrized over some coordinate system as $x^\mu(\lambda)$, marked in blue. It's tangent vector at some point is marked in red, labelled v^μ.

Consider a worldline through a spacetime manifold M that is parametrized using a particular coordinate system such that we can label different points along the worldline using the coordinates as functions of the parametrized variable such that

$$x^\mu \to x^\mu(\lambda)$$

We would like to find out the path such that we know how the coordinates x^μ depend on λ. For this, we use a **tangent vector** or **contravariant vector** to the worldline

$$\boxed{v^\mu = \frac{\partial x^\mu}{\partial \lambda}} \tag{2.4}$$

We know that given some new coordinates x^ν, we will get the tangent vector in new coordinates as

$$v^\nu = \frac{\partial x^\nu}{\partial \lambda}$$

If we'd like to obtain the tangent vector v^ν in terms of the coordinates x^μ, we recall the chain rule:

$$\boxed{v^\nu = \left(\frac{\partial x^\nu}{\partial x^\mu}\right)\frac{\partial x^\mu}{\partial \lambda}} \tag{2.5}$$

it is important to remember that ν is a free index whereas μ is a dummy index for a sum from 0 to 3 such that we have 4 equations

$$v^\nu = \sum_{\mu=0}^{3}\left(\frac{\partial x^\nu}{\partial x^\mu}\right)\frac{\partial x^\mu}{\partial \lambda} \quad \text{for } \nu = 0, 1, 2, 3$$

This is the same as equations (2.5) except those are in Einstein summation notation, while these ones include the summation. It is evident that we are applying the chain rule to the tangent vector v^μ with the coordinate's Jacobian $J = \frac{\partial x^\nu}{\partial x^\mu}$ to obtain a tangent vector in new coordinates v^μ.

Note. Despite x^μ having a contravariant index, we have to remember that we are talking about coordinates and not vectors. Notice how the basis vectors e_μ for the vector $\vec{v}$ in the coordinates x^μ also have a covariant index despite vectors having no indices. This is because the tensor notation is saved for when we refer specifically to tensors, or otherwise noted. For example, the tangential vector v^i is indeed a contravariant component as it will be transformed by the inverse transformation that yields a change of basis, in this case, the parametrization from some coordinates onto a worldline. Hence, we can use other notations as long as they are clearly understood in the context in which they are dealt with.

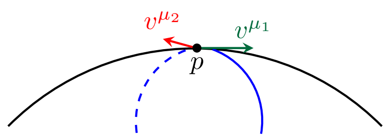

Figure 7: An arbitrary 3D curved manifold with a point. A single curve is shown crossing the point along the manifold in blue. Consider the black curvature over the manifold as another curve crossing the point over the manifold. The tangent vector v^{μ_1} corresponds to the black curve whereas v^{μ_2} corresponds to the blue curve.

Consider some curved manifold M such that we can parametrize curves that go through some point p on M. To define a tangent vector at such a point, recall equation (2.4):

$$v^\mu = \frac{\partial x^\mu}{\partial \lambda}$$

We would not like to extend the notion of a tangent vector to all possible worldlines on an arbitrary manifold given a point such that we may obtain a set of all possible tangent vectors represented at this point.

Definition 2.11. The **tangent space** for some point p in a D-dimensional manifold M is a vector space denoted as the set of all tangent vectors spanned by all the possible curves that intersect $p \in M$, and is denoted $T_p(M)$.

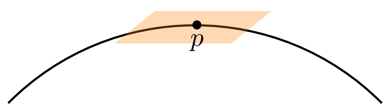

Figure 8: A tangent space represented as a plane on a point over some 3D manifold.

A tangent vector as we have defined is an example of a vector that is defined for every point on a curve in a manifold. More generally, a tangent vector is a vector that is attached to every point $p \in M$.

Definition 2.12. A **vector field** is the set of all vectors defined for every point in a manifold M. As such, a vector field at some point $p \in M$ is itself a vector.

We can think of vectors in many ways, particularly in geometric ones. For instance, vectors can represent differential operators.

Definition 2.13. Consider a scalar function over some coordinates given by $a(x^\mu)$ and a tangent vector over the same coordinates $v^\mu(x^\mu)$. The **differential operator**, otherwise known as the **directional derivative** v is a scalar function such that

$$\boxed{va = v^\mu \frac{\partial}{\partial x^\mu} a = v^\mu a_\mu} \tag{2.6}$$

Clearly, we have that $a_\mu = \frac{\partial a}{\partial x^\mu} = \partial_\mu a$. It may also be clear that the we are yielding a scalar, or a (0,0)-tensor. To see this, let us write v using the chain rule for a change of coordinates.

$$\begin{aligned} va(x^\mu) &= v^{\mu'} \frac{\partial a(x^\mu)}{\partial x^{\mu'}} \\ &= v^\mu \left(\frac{\partial x^{\mu'}}{\partial x^\mu}\right)\left(\frac{\partial x^\nu}{\partial x^{\mu'}}\right)\frac{\partial a(x^\mu)}{\partial x^\nu} \\ &= \left(\sum_{\mu=0}^{D-1} v^\mu \left(\frac{\partial x^{\mu'}}{\partial x^\mu}\right)\right)\left(\sum_{\nu=0}^{D-1}\left(\frac{\partial x^\nu}{\partial x^{\mu'}}\right)\frac{\partial a(x^\mu)}{\partial x^\nu}\right) \end{aligned}$$

We have two different Jacobians: The first $\frac{\partial x^{\mu'}}{\partial x^{\mu}}$ is for changing from x^{μ} to $x^{\mu'}$ coordinates, and the second $\frac{\partial x^{\mu}}{\partial x^{\mu'}}$ is for changing from $x^{\mu'}$ to x^{μ} coordinates. They each are respective inverses of the other, and multiplied they yield an identity matrix.

Definition 2.14. The **Kronecker delta** is defined as a (1,1)-tensor yielded by the multiplication of a Jacobian by it's own inverse, namely

$$\boxed{\delta^{\nu}_{\mu} = \frac{\partial x^{\mu'}}{\partial x^{\mu}} \frac{\partial x^{\nu}}{\partial x^{\mu'}} = \begin{cases} 1, & \text{if } \mu = \nu \\ 0, & \text{if } \mu \neq \nu \end{cases}} \tag{2.7}$$

The fact that a Jacobian times its inverse results in the Kronecker delta implies that the differential operator on a scalar function becomes

$$va = v^{\mu} \delta^{\nu}_{\mu} \frac{\partial a}{\partial x^{\mu}} = v^{\mu} \frac{\partial a}{\partial x^{\mu}}$$

Clearly, this means that this differential operator v transforms a scalar to a scalar. More explicitly, it transforms a (0,0)-Tensor into another (0,0)-Tensor. In calculus terms, it yields the directional derivative of a scalar in the direction of v.

Example 2.7. A scalar function on a tangent vector

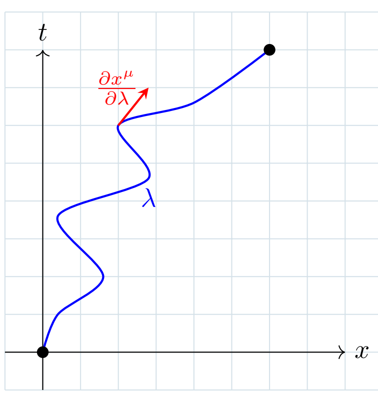

Figure 9: A worldline on a spacetime manifold parametrized over some coordinate system as $x^{\mu}(\lambda)$, marked in blue. It's tangent vector at some point is marked in red, labelled $\frac{\partial x^{\mu}}{\partial \lambda}$.

Considering a worldline over some manifold M, we define a scalar function over the coordinates such that

$$a\big(x^{\mu}(\lambda)\big) = a(\lambda)$$

By the chain rule, the directional derivative on the scalar a is hence given by

$$\begin{aligned} \frac{da}{d\lambda} &= \frac{\partial x^{\mu}}{\partial \lambda} \frac{\partial a}{\partial x^{\mu}} \\ &= v^{\mu} \frac{\partial}{\partial x^{\mu}} a \\ &= va \end{aligned}$$

This leads us to see that the differential operator v takes derivatives along a curve to which the vector v^{μ} is a tangent vector. Sometimes, we might think of ∂_{μ} as a *basis* for the tangent vectors v^{μ}. Thus, a directional derivative is neatly given by a tangential component and a basis component:

$$v = v^{\mu} \partial_{\mu}$$

Exercise 2.4. Consider the manifold $M = \mathbb{R}^2$ in both Cartesian (x, y) and Polar coordinates (r, θ) such that a tangential vector in Cartesian coordinates is given by

$$v^\mu = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

(a) Write the tangential vector v^μ in polar coordinates.

(b) Write the differential operator $\frac{\partial}{\partial x^\mu}$ in Polar coordinates.

(c) Explain how (a) and (b) relate to each other.

Theorem 2.8. *The tangential vectors v^μ of a vector $\vec{v}$ on some basis $\frac{\partial}{\partial x^\mu}$ transform from coordinates x^μ to coordinates $x^{\mu'}$ as*

$$v^\mu \to v^{\mu'} = \frac{\partial x^{\mu'}}{\partial x^\mu} v^\mu$$

Where $x^{\mu'} = x^{\mu'}(x^\mu)$ and $\frac{\partial x^{\mu'}}{\partial x^\mu}$ is the Jacobian from x^μ to $x^{\mu'}$.

We can therefore represent tangential vectors v^μ in a coordinate independent way as a differential operator which takes scalars to scalars. Now, let's introduce our next tensor.

Definition 2.15. Given a D-dimensional manifold M, a **one-form** is a (0,1)-tensor given as a set of D functions on some coordinates x^μ denoted by w_μ. It is otherwise known as **covariant vector**.

Theorem 2.9. *A one form w_μ on some coordinates x^μ transforms as*

$$w_\mu \to w_{\mu'} = \frac{\partial x^\mu}{\partial x^{\mu'}} w_\mu$$

Corollary 2.10. *Given two coordinates x^μ and $x^{\mu'}$ on a D-dimensional manifold M, the Jacobian $\frac{\partial x^\nu}{\partial x^{\mu'}}$ that transforms a one form as $w_\nu \to w_{\mu'} = \frac{\partial x^\nu}{\partial x^{\mu'}} w_\nu$ is the inverse of the Jacobian $\frac{\partial x^{\mu'}}{\partial x^\mu}$ that transforms the components of a vector as $v^\mu \to v^{\mu'} = \frac{\partial x^{\mu'}}{\partial x^\mu} v^\mu$ such that*

$$\frac{\partial x^{\mu'}}{\partial x^\mu} \frac{\partial x^\nu}{\partial x^{\mu'}} = \delta^\nu_\mu$$

The difference between a vector and a one form is that a vector can be represented by contravariant components v^μ whereas a one-form is represented by covariant components w_μ. Moreover, this extends to why the Jacobians are written in each way respectively for a one form w_μ and a vector v^μ, as we may get a scalar by summing the multiplication of each function as $v^\mu w_\mu = a$. To show this, let us change from coordinates x^μ to coordinates $x^{\mu'}$ with the following transformation

$$v^\mu w_\mu \to v^{\mu'} w_{\mu'} = \left(v^\mu \frac{\partial x^{\mu'}}{\partial x^\mu} \right) \left(\frac{\partial x^\nu}{\partial x^{\mu'}} w_\nu \right)$$

By Corollary 2.10, we have

$$v^{\mu'} w_{\mu'} = v^{\mu} \delta^{\nu}_{\mu} w_{\nu} = v^{\mu} w_{\mu}$$

Definition 2.16. The product of a one-form w_μ and a vector v^μ is known as a **dot-product**.

So formally, we can think that a one form w_μ is a linear mapping from vector components v^μ to scalars a. For this very reason, sometimes one-forms can be referred to as *dual-vectors* that mirror vectors. In the same way a vector belongs to some tangent space, a one form also belongs to a *dual-tangent space* or *cotangent space* denoted $T^*_p(M)$. However, this language is saved for much more mathematical formulations of General Relativity. The mention of this vocabulary is done for the sake of reference for further insight into other texts.

In linear algebra, we usually represent vectors as column vectors and one-forms as row vectors. In particular, considering a D-dimensional manifold M we have

$$w_\mu = \begin{pmatrix} w_0 & w_1 & \dots & w_{D-1} \end{pmatrix} \quad v^\mu = \begin{pmatrix} v^1 \\ v^2 \\ \vdots \\ v^{D-1} \end{pmatrix} \implies w_\mu v^\mu = \sum_{\nu=0}^{D-1} w_\nu v^\nu$$

Whether a tensor has a covariant (lower) or contravariant (upper) index will impact how that specific tensor transforms under a coordinate change. If we instead have $v^\mu w_\mu$ we would get a 3×3 matrix but would only sum the trace of this matrix, yielding the same result as $w_\mu v^\mu$.

Definition 2.17. Whenever we set a covariant and a contravariant index as equal in a tensor and subsequently sum over them, we denote that as **contraction of indices**.

Theorem 2.11. *For some scalar function $a = a(x^\mu)$ on a D-dimensional manifold M given some coordinates x^μ, we can define a one-form w_μ whose componets are given by*

$$\boxed{\frac{\partial a}{\partial x^\mu} = \partial_\mu a} \tag{2.8}$$

Now let us change the components of a one-form w_μ from coordinates x^μ to coordinates $x^{\mu'}$, we use the chain rule to get

$$\partial_\mu a \to \partial_{\mu'} a = \frac{\partial x^\mu}{\partial x^{\mu'}} \frac{\partial a(x^\mu)}{\partial x^\mu}$$

And just like we write vectors using a basis independent way $\vec{v} = v^\mu \partial_\mu$, we can write a one-form in a basis independent way such that

$$w = w_\mu dx^\mu$$

Definition 2.18. We denote dx^μ as an **infinitesimally small displacement of coordinates** x^μ.

Given another set of coordinates $x^{\mu'}$ we can obtain a one form in x^μ coordinates as

$$w = w_\mu dx^\mu = \left(w_{\mu'} \frac{\partial x^{\mu'}}{\partial x^\mu} \right) \left(\frac{\partial x^\mu}{\partial x^{\nu'}} dx^{\nu'} \right)$$

By Corollary 2.10, we again have $\delta^{\mu'}_{\mu}$, which gives

$$w = w_\mu dx^\mu = w_{\mu'} dx^{\mu'}$$

Now, let us go back to Theorem 2.11 from which we see that for some coordinates x^μ, there exists a special notation for the one-form $w_\mu = \partial_\mu a$, whose components are the derivatives of some scalar function $a = a(x^\mu)$. This notation is called **integration** of a and is denoted

$$\boxed{da = \partial_\mu a \, dx^\mu} \tag{2.9}$$

Example 2.8. One-form of a Cartesian manifold

Consider the $\mathbb{R}^2$ manifold. Let us define a Cartesian coordinate system such that we can define the scalar functions

$$\begin{aligned} a(x^1) &= x^1 = x \\ a(x^2) &= x^2 = y \end{aligned}$$

We can easily obtain the components of the one-form:

$$\begin{aligned} \partial_1 a(x^1) &= \begin{pmatrix} 1 & 0 \end{pmatrix} \\ \partial_2 a(x^2) &= \begin{pmatrix} 0 & 1 \end{pmatrix} \end{aligned}$$

Therefore, the infinitesimal displacement for each respective component in the one-form will be given by

$$\begin{aligned} \partial_1 a(x^1) dx^1 &= \begin{pmatrix} dx & 0 \end{pmatrix} \\ \partial_2 a(x^2) dx^2 &= \begin{pmatrix} 0 & dy \end{pmatrix} \end{aligned}$$

The one-form can be more easily explained as the sum of the derivatives of each coordinate. It yields an independent basis in the same way that the basis for a vector $\frac{\partial}{\partial x^\mu} = \partial_\mu$ does. We can say as physicists that a one-form's one and only purpose is to be integrated such that

$$\int da$$

Theorem 2.12. *The **Fundamental Theorem of Calculus** states that, given two points p and q defined on a function f, it follows that*

$$\boxed{\int_p^q df = f(q) - f(p)} \tag{2.10}$$

We now have enough insight into some basic tensors to generalize and define it formally. Definition 2.6 does provide us with some useful information, but we had to define more things along the way in order to understand what we meant better. Hence, now let us define it using everything we've learned thus far.

Definition 2.19. A (k, ℓ)-tensor is a set of $D^{k+\ell}$ functions denoted by

$$T^{\mu_1,\ldots,\mu_k}_{\nu_1,\ldots,\nu_\ell} = \left(\frac{\partial x^{\mu_{1'}}}{\partial x^{\mu_1}}\right)\cdots\left(\frac{\partial x^{\mu'_k}}{\partial x^{\mu_k}}\right)\left(\frac{\partial x^{\nu_1}}{\partial x^{\nu_{1'}}}\right)\cdots\left(\frac{\partial x^{\nu_\ell}}{\partial x^{\nu_{\ell'}}}\right)$$

This looks quite complicated. So we need only remember index contraction such that this eventually simplified. It therefore extremely important to note the ordering and placement of the indices, as changing either will inevitably change the entire tensor. On a last note, we can also write a tensor using a tensor product. By use of the tensor product $T \otimes T'$, a tensor becomes much easier to write down in a coordinate invariant way via contraction of indices.

$$T = T^{\mu_1,\ldots,\mu_k}_{\nu_1,\ldots,\nu_\ell}\ \partial_{\mu_1} \otimes \cdots \otimes \partial_{\mu_k} \otimes dx^{\nu_1} \otimes \cdots \otimes dx^{\nu_\ell}$$

Let us clean this up further. We clearly see that this involves the outer product of several vector spaces, whose elements are vectors, and several dual vector spaces, whose elements are one-forms. Hence, let us define V_{μ_i} as the vector space for the basis ∂_{μ_i} where $i = 1, \ldots, k$ and let us define the dual vector space W^{ν_j} for the infinitesimal displacement dx^{ν_j} where $j = 1, \ldots, \ell$. We can then say that

$$T = T^{\mu_1,\ldots,\mu_k}_{\nu_1,\ldots,\nu_\ell} = V^{\mu_1} \otimes \cdots \otimes V^{\mu_k} \otimes W_{\nu_1} \otimes \cdots \otimes W_{\nu_\ell}$$

Moreover, let us give a special notation for the outer product of several vector spaces:

$$V^1 \otimes \cdots \otimes V^n = V^{\otimes_n}$$

Finally, a tensor can be written in terms of vector spaces and dual vector spaces as

$$T = T^{\mu_1,\ldots,\mu_k}_{\nu_1,\ldots,\nu_\ell} = V^{\otimes_{\mu_k}} \otimes W_{\otimes_{\nu_\ell}}$$

Example 2.9. The metric tensor

Consider the (0,2)-tensor that denotes the variations of a spacetime metric $g_{\mu\nu}$. We can write this tensor as

$$\boxed{g = g_{\mu\nu} dx^\mu \otimes dx^\nu = g_{\mu\nu} dx^\mu dx^\nu} \tag{2.11}$$

2.3 Tensor Algebra

We have the following operations on tensors:

- **Addition**

  Two tensors of the same weight $A^{\mu_1,\dots,\mu_k}_{\nu_1,\dots,\nu_\ell}$ and $B^{\mu_1,\dots,\mu_k}_{\nu_1,\dots,\nu_\ell}$ can be added to obtain another tensor of the same weight $C^{\mu_1,\dots,\mu_k}_{\nu_1,\dots,\nu_\ell}$. To achieve this, we simply add the components of each vector correspondingly to the indices such that we get

$$A^{\mu_1,\dots,\mu_k}_{\nu_1,\dots,\nu_\ell} + B^{\mu_1,\dots,\mu_k}_{\nu_1,\dots,\nu_\ell} = C^{\mu_1,\dots,\mu_k}_{\nu_1,\dots,\nu_\ell}$$

- **Multiplication**

  We can multiply a (k,ℓ)-tensor $A^{\mu_1,\dots,\mu_k}_{\nu_1,\dots,\nu_\ell}$ with a (k',ℓ')-tensor $B^{\mu_1,\dots,\mu_{k'}}_{\nu_1,\dots,\nu_{\ell'}}$ to get another $(k+k',\ell+\ell')$-tensor $C^{\mu_1,\dots,\mu_{k+k'}}_{\nu_1,\dots,\nu_{\ell+\ell'}}$. So a tensor transforms as a tensor under multiplication, namely

$$A^{\mu_1,\dots,\mu_k}_{\nu_1,\dots,\nu_\ell} \otimes B^{\mu_1,\dots,\mu_{k'}}_{\nu_1,\dots,\nu_{\ell'}} = C^{\mu_1,\dots,\mu_{k+k'}}_{\nu_1,\dots,\nu_{\ell+\ell'}}$$

- **Contraction**

  Given a (k,ℓ)-tensor $A^{\mu_1,\dots,\mu_k}_{\nu_1,\dots,\nu_\ell}$, we can contract covariant and contravariant indices to get a $(k-1,\ell-1)$-tensor $B^{\mu_1,\dots,\mu_{k-1}}_{\nu_1,\dots,\nu_{\ell-1}}$ For instance, $v^\mu w_\nu$ is a (1,1)-tensor while $v^\mu w_\mu$ is a $(0,0)$-tensor.

Exercise 2.5. Consider the manifold $\mathbb{R}^2$ in coordinates $x^\mu = (x,y)$ and $x^{\mu'} = (r,\theta)$. Given the following tangent vector and the one form in x^μ coordinates

$$v^\mu = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad w_\mu = \begin{pmatrix} 1 & 0 \end{pmatrix}$$

(a) Find the tangent vector $v^{\mu'}$ and the one-form $w_{\mu'}$ in $x^{\mu'}$ coordinates.

(b) Represent $v = v^\mu \partial_\mu$ and $w = w_\mu dx^\mu$ in coordinate-free notation.