Introduction to General Relativity for Enthusiasts

Zihan Yan

Michaelmas 2019

Preface

Why General Relativity?

Newtonian mechanics is useful for our everyday life, the most common motions we see around us are low in velocity (compared with the speed of light) and occur in a rather weak and approximately uniform gravitational field. But, generally, is it good enough to give us the right physics picture and understanding? The answer is no. Problems arose everywhere: Maxwell's equations suggests that the speed of light in vacuum is independent of observers, which is not compatible with Newtonian mechanics where velocity transformations for different observers are just simple addition/subtraction; Michelson-Morley experiment (1887) supported the above prediction, disproving Newtonian mechanics. This leads to the formulation of Special Relativity. But, when considering (Newtonian) gravity within the framework of SR, it failed since the gravitation in Newtonian picture is assumed as an instantaneous action at a distance which is not allowed in SR (no speed can exceed the speed of light). Also, Newtonian gravitation itself has flaws such as its failure in predict/explain the perihelion precession of Mercury and the bending of light. Thus, to incorporate SR and gravitation, General Relativity was born. I would say it is one of the greatest achievement of human mind and human civilisation obviously by its elegance and accuracy.

How do we understand nature through Mathematics?

To me, I believe that Mathematics is *a priori*, meaning that certain mathematical structures already exist before we discover them, which is knowledge acquired independent of our experience. Also, the universality of Mathematics is reflected in the fact that for *an average sane man*, the same set of preconditions, through Mathematical procedures, always leads to the same set of conclusions, which must be agreed by different *sane* people. Thus, by constructing some common well-chosen preconditions (may be known as axioms, postulates, laws, etc.), we can derive a self-consistent mathematical structure that can give the True/False value of a certain set of statements. If we construct the mathematical structure in such a way that is isomorphic to the laws of nature to some accuracy, it gives a working theory using which we can predict certain natural phenomena. We may never "understand" the true underlying laws of nature, but we indeed understand the Mathematics we built (it is a subset of human mind!) that resembles the nature.

Contents

1 Special Relativity from an elementary perspective

In this section, I will show how to derive some of the most important results in Special Relativity in a bottom-up fashion. We would do the same thing again in a top-down manner by the time we have equipped better mathematical tools. (In section 5 and 6.)

1.1 Some Fundamental Definitions

Definition 1.1 (Spacetime). Spacetime is some mathematical model that fuses three dimensions of space and one dimension of time into a single four-dimensional continuum.

The above definition is based on the genuine observation that space is three-dimensional and time is one-dimensional.

Definition 1.2 (Event). An event is a physical situation or occurrence associated with a certain point in spacetime.

Definition 1.3 (Frame of reference). A frame of reference is a set of coordinates $(t, x^i), i \in \{1, 2, 3\}$ that standardises the measurement in the spacetime with respect to a certain observer. t is the temporal coordinate according to a system of synchronised clocks at rest with the observer; x^i are the spatial coordinates according to the observer. We can label any event by (t, x^i).

Definition 1.4 (Inertial frame). An inertial frame is a frame of reference where Newton's first law holds, i.e. for free particles (total external force is zero), the trajectory $x^i(t)$ satisfies

$$\frac{\mathrm{d}^2 x^i}{\mathrm{d}t^2} = 0$$

1.2 A Review of Newtonian View of Spacetime

Postulate (Newtonian Spacetime). *In Newtonian Mechanics, the space and time are independent and are absolute, i.e. the spatial distance of any two points is invariant and the time is synchronised for any observer.*

Claim 1.1. *In Newtonian spacetime, any two inertial frames S, S' differ by a translation and/or a rotation and/or a relative motion at constant velocity.*

Proof. S and S' have coordinate systems (t, x^i) and (t', x'^i), respectively. Since time is absolute, $t' = t$. For some free particle trajectory $x'^i(t')$ and $x^i(t)$, we have

$$\frac{\mathrm{d}^2 x'^i}{\mathrm{d}t'^2} = \frac{\mathrm{d}^2 x^i}{\mathrm{d}t^2} = 0$$

Then

$$\begin{aligned}\frac{\mathrm{d}^2 x'^i}{\mathrm{d}t'^2} &= \frac{\mathrm{d}}{\mathrm{d}t'}\left(\frac{\mathrm{d}x'^i}{\mathrm{d}t'}\right) = \frac{\mathrm{d}}{\mathrm{d}t'}\left(\frac{\mathrm{d}t}{\mathrm{d}t'}\frac{\mathrm{d}x^j}{\mathrm{d}t}\frac{\partial x'^i}{\partial x^j}\right) \\ &= \underbrace{\frac{\mathrm{d}^2 t}{\mathrm{d}t'^2}}_{=0}\frac{\mathrm{d}x^j}{\mathrm{d}t}\frac{\partial x'^i}{\partial x^j} + \left(\frac{\mathrm{d}t}{\mathrm{d}t'}\right)^2 \underbrace{\frac{\mathrm{d}^2 x^j}{\mathrm{d}t^2}}_{=0}\frac{\partial x'^i}{\partial x^j} + \left(\frac{\mathrm{d}t}{\mathrm{d}t'}\right)^2\left(\frac{\mathrm{d}x^j}{\mathrm{d}t}\right)^2\frac{\partial^2 x'^i}{\partial {x^j}^2} = 0\end{aligned}$$

Note that the Einstein summation convention is used throughout the text, unless otherwise stated.

Therefore we deduce

$$\frac{\partial^2 x'^i}{\partial x^{j\,2}} = 0$$

We have

$$x'^i = a^i{}_j x^j + c^i(t)$$

where $a^i{}_j$ are constants, $c^i(t)$ is some function in time. As $\frac{\mathrm{d}^2 x'^i}{\mathrm{d}t^2} = 0$, we have

$$c^i(t) = v^i t + d^i$$

for v^i, d^i constants. Then we have the most general transformational law as

$$\boxed{x'^i = a^i{}_j x^j + v^i t + d^i}$$

Since $a^i{}_j$ satisfies (by absolute space)

$$\delta_{kl}\mathrm{d}x^k\mathrm{d}x^l = \delta_{ij}a^i{}_k a^j{}_l \mathrm{d}x^k\mathrm{d}x^l$$

where $\delta_{ij} = \mathrm{diag}(1,1,1)$, it is orthogonal since for $\mathbf{A} = (a^i{}_j)$,

$$\delta_{kl} = \delta_{ij}a^i{}_k a^j{}_l \qquad \Leftrightarrow \qquad \mathbf{A}^T\mathbf{A} = \mathbf{I}$$

and can be interpreted as rotation. □

1.3 Motivation of Special Relativity

Let us discuss the prediction on speed of light by Maxwell's equations, mentioned in the Preface. The Maxwell's equations are

$$\begin{aligned}
\boldsymbol{\nabla}\cdot\mathbf{E} &= \rho/\epsilon_0\\
\boldsymbol{\nabla}\cdot\mathbf{B} &= 0\\
\boldsymbol{\nabla}\times\mathbf{E} &= -\dot{\mathbf{B}}\\
\boldsymbol{\nabla}\times\mathbf{B} &= \mu_0\mathbf{J} + \mu_0\epsilon_0\dot{\mathbf{E}}
\end{aligned}$$

where $\mathbf{E}, \mathbf{B}, \rho, \mathbf{J}$ are electric field, magnetic flux density, charge density and current density, respectively. ϵ_0 is the electric permittivity of vacuum, μ_0 is the magnetic permeability of vacuum, both are constants.

In vacuum,

$$\begin{aligned}
\boldsymbol{\nabla}\cdot\mathbf{E} &= 0\\
\boldsymbol{\nabla}\cdot\mathbf{B} &= 0\\
\boldsymbol{\nabla}\times\mathbf{E} &= -\dot{\mathbf{B}}\\
\boldsymbol{\nabla}\times\mathbf{B} &= \mu_0\epsilon_0\dot{\mathbf{E}}
\end{aligned}$$

We will shortly see that the Maxwell's equations in vacuum predict wave solutions of $\mathbf{E}$ and $\mathbf{B}$, simply known now as the electromagnetic waves. Take the curl of $\boldsymbol{\nabla}\times\mathbf{E}$,

$$\boldsymbol{\nabla}\times(\boldsymbol{\nabla}\times\mathbf{E}) = -\boldsymbol{\nabla}\times\dot{\mathbf{B}} = -\mu_0\epsilon_0\ddot{\mathbf{E}}$$

Also, by

$$\boldsymbol{\nabla} \times (\boldsymbol{\nabla} \times \mathbf{E}) = \boldsymbol{\nabla}(\boldsymbol{\nabla} \cdot \mathbf{E}) - \nabla^2 \mathbf{E}$$

we have (and reproduce the similar procedure for $\mathbf{B}$)

$$\nabla^2 \mathbf{E} = \mu_0 \epsilon_0 \ddot{\mathbf{E}} \qquad \nabla^2 \mathbf{B} = \mu_0 \epsilon_0 \ddot{\mathbf{B}} \tag{1}$$

which are wave equations

$$\nabla^2 \mathbf{E} = \frac{1}{c^2} \ddot{\mathbf{E}} \qquad \nabla^2 \mathbf{B} = \frac{1}{c^2} \ddot{\mathbf{B}}$$

with wave speed

$$c = \frac{1}{\sqrt{\mu_0 \epsilon_0}} \tag{2}$$

This suggests that in vacuum the speed of electromagnetic waves is c, a constant with respect to any observer since μ_0, ϵ_0 are constants. This was supported by the Michelson-Morley experiment.

1.4 Postulates of Special Relativity

POSTULATE 1. *(Principle of Relativity) All laws of physics are the same in all inertial frames.*

POSTULATE 2. *The speed of light (in vacuum) is the same in all inertial frames.*

The elegance of Special Relativity is that we can derive the whole theory (no matter top-down or bottom-up) merely from these two postulates.

In Special Relativity, we use coordinates (ct, x, y, z) to denote events. This makes the dimension of all coordinates the same (length).

1.5 An elementary approach: thought experiments

Here, we use an elementary approach to construct the building blocks of Special Relativity bottom up. We can consider the physics under the two postulates using thought experiments on some generic examples.

1.5.1 Time dilation

We consider a train of height h moving at constant speed v with respect to frame S. The co-moving frame with the train is denoted as S'. In S' a ray of light was emitted vertically from point O, we denote this event $\mathcal{A}$. Then the ray was reflected by a mirror back to O, we denote this event $\mathcal{B}$.

Note that the speed of light c is the same for both frames.

In S', the time **measured by the co-moving observer** between $\mathcal{A}, \mathcal{B}$ is $t' = 2h/c$.

In S, the path in which the light ray travels seen by **local observer** have length $2\sqrt{(vt/2)^2 + h^2}$, where t is the time between $\mathcal{A}, \mathcal{B}$ by the observer in S. Then we have

$$ct = 2\sqrt{\left(\frac{vt}{2}\right)^2 + h^2}$$

Rearrange, we have the famous time dilation formula

$$\boxed{t = \gamma t'} \tag{3}$$

where

$$\gamma = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}$$

is the Lorentz factor.

This means that the flow of time in S' **seen by** S is slower, although observers in S' wouldn't feel any difference in their own sense of time. Relatively, the time in S seen by S' is also dilated.

Note. Note, for Lorentz factor to be physical, the magnitude of velocity v of any (massive, will be shown later) object cannot reach or exceed the speed of light c.

1.5.2 Length contraction

Simultaneity is relative. This can be seen by a simple thought experiment. Consider a train moving at a constant speed v with respect to a platform. Say, the length of the moving train coincide with the length of the platform. When the ends of the train align with the ends of the platform, we emit light rays from both ends on the platform back to the mid-point. The observer at the mid-point of the platform sees the light ray arrive at the point simultaneously, while, on the moving train, the observer at the mid-point of the train sees one ray arrives at the point earlier than the other, simply because postulate 2.

Definition 1.5 (Length measurement). A length measurement is the measurement of spatial distance at the same time.

This definition seems subtle, but it would be clearer later.

To investigate length contraction, we need yet another thought experiment. Since measurements of the length of moving objects are generally hard for a macroscopic observer since the mechanism that reflects the *simultaneous* measurement of both ends is complicated to think about. Instead, we use a genuine point-like observer so that itself can be a marker of length.

Consider a rod rest in frame S, with length l_0. There is a point-like observer moving at constant speed v along the direction of the rod, whose frame is denoted as S'. The observer passes the two ends of the rod with distance measured by it l', taking time t' in S'. The time for observer in S to witness the moving observer to pass both ends is t, by time dilation, $t = \gamma t'$. But since the relative speed should agree in both frames, we have

$$l' = vt' \qquad \text{and} \qquad l_0 = vt$$

This simply gives the famous length contraction formula:

$$\boxed{l' = \frac{l_0}{\gamma}} \tag{4}$$

This means that moving object appear to have shorter length (with respect to an observer at rest) in the direction of motion.

NOTE. The subtlety here is: what happens to the lengths perpendicular to the direction of motion? The answer is they are unchanged since the simultaneity of this measurement agrees in both frame. (Think about a measurement using light rays.)

1.5.3 (Naïvely) Lorentz Transformation

From the above results, we can "naïvely" derive the Lorentz Transformation.

> DEFINITION 1.6 (Standard configuration). Two frames S and S' are said to be in standard configuration if their spacetime origin (0,0,0,0) coincide.

Take a frame S to be rest. Without loss of generality, we can take another frame S' in standard configuration with S with all spatial axes parallel to S, moving at constant speed v along x-direction with respect to S. Consider two events $\mathcal{A}_{S'} = (ct', 0, 0, 0)$ and $\mathcal{B}_{S'} = (ct', x', 0, 0)$ viewed in S'. In S, these are $\mathcal{A}_S = (ct, vt, 0, 0)$ and $\mathcal{B}_S = (ct, x, 0, 0)$ respectively. By length contraction, we have

$$x' = \gamma(x - vt) \tag{5}$$

Similarly, we take $\mathcal{C}_S = (ct, 0, 0, 0)$ and $\mathcal{D}_S = (ct, x, 0, 0)$, by the same argument we find the counterpart of above

$$x = \gamma(x' + vt') \tag{6}$$

Substitute one in the other, we have

$$ct' = \gamma\left(ct - \frac{vx}{c}\right) \qquad \text{and} \qquad ct = \gamma\left(ct' + \frac{vx'}{c}\right) \tag{7}$$

Note that $y' = y, z' = z$ since these axes are perpendicular to the direction of motion. This can be written compactly as

$$\begin{pmatrix} ct' \\ x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} \gamma & -\gamma\beta & 0 & 0 \\ -\gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix} \tag{8}$$

where $\beta = v/c$.

This is the Lorentz Transformation, the way we link two inertial frames with relative motion in Special Relativity.

1.6 Spacetime perspective

EXERCISE. Verify that, under Lorentz transformation,

$$c^2\Delta t'^2 - \Delta x'^2 - \Delta y'^2 - \Delta z'^2 = c^2\Delta t^2 - \Delta x^2 - \Delta y^2 - \Delta z^2$$

DEFINITION 1.7 (Spacetime interval). Infinitesimally, we define (Minkowski) spacetime interval

$$\mathrm{d}s^2 = c^2\mathrm{d}t^2 - \mathrm{d}x^2 - \mathrm{d}y^2 - \mathrm{d}z^2 \tag{9}$$

which is invariant with respect to different observers.

We will encounter a more general form of this, known as *metric* later when we discuss Differential Geometry. The spacetime with the above metric is called a Minkowski spacetime. Generally, there can be different metric on spacetime, this leads to the formulation of General Relativity.

DEFINITION 1.8 (Proper time). The proper time τ is defined as

$$\mathrm{d}\tau^2 = \frac{1}{c^2}\mathrm{d}s^2$$

To see some interesting property of Minkowski spacetime, we have the following construction.

DEFINITION 1.9 (Rapidity). The rapidity ψ is defined such that $\beta = \tanh\psi$.

Then we easily have

$$\gamma = \cosh\psi \qquad \text{and} \qquad \gamma\beta = \sinh\psi$$

And

$$\begin{pmatrix} ct' \\ x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} \cosh\psi & -\sinh\psi & 0 & 0 \\ -\sinh\psi & \cosh\psi & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix} \tag{10}$$

It seems like a hyperbolic "rotation", called a *boost*.

Also, the significance of rapidity can be seen from the fact that $v \to \pm c \Leftrightarrow \psi \to \pm\infty$. Later, we will see the constant acceleration is in fact linear in time for rapidity, making calculations more convenient in terms of rapidity.

1.7 Spacetime Diagrams and Light Cone Structure

DEFINITION 1.10 (Properties of spacetime interval). The spacetime interval $\mathrm{d}s^2 > 0$ is called timelike; $\mathrm{d}s^2 = 0$ is called null or lightlike; $\mathrm{d}s^2 < 0$ is called spacelike.

DEFINITION 1.11 (Light cone). The light cone of some event $\mathcal{A}$ is the set of all events that can be connected to $\mathcal{A}$ by light signals.

Definition 1.12 (Causal future/past). The causal future of some event $\mathcal{A}$ is the set of all events within the forward light cone of $\mathcal{A}$; the causal past of $\mathcal{A}$ is the set of all events within the past light cone of $\mathcal{A}$.

We now introduce the spacetime diagram in 2D spacetime. Spacetime diagram is useful when consider how an event is observed in different frames.

Let S be a rest frame, S' moves at constant speed v at x-direction with respect to S. Then by Lorentz Transformation, we have

$$x' = \gamma(x - \beta ct) \qquad \text{and} \qquad ct' = \gamma(ct - \beta x)$$

For $x' = 0$ we have $x = vt$; for $ct' = 0$ we have $x = t/v$. These two lines define the axes of S' seen in S: $x = vt \Leftrightarrow ct'$-axis, $x = t/v \Leftrightarrow x'$-axis. And it is easy to see the angle difference between x' and x axes (the same for ct' and ct) α satisfies $\tan\alpha = \beta$.

But the scales on the axes are not apparent at first sight. Consider a family of curves

$$(ct)^2 - x^2 = (ct')^2 - (x')^2 = \pm s_0^2$$

for specified constant s_0, we obtain hyperbolae for constant spacetime interval, from which we can label the axes scale on the spacetime diagram. Of course, these scales are non-uniform along each axes.

Example (Doppler effect). Consider an observer at rest at spatial origin of frame S and an EM-wave emitter in frame S' moving at constant velocity v in x-direction with respect to S away from the observer. Let $\mathcal{A}, \mathcal{B}$ denote two consecutive events where the emitter send a signal. $\mathcal{C}, \mathcal{D}$ denote two consecutive events where the observer receives a signal. By time dilation

$$t_{\mathcal{AB}} = \gamma t'_{\mathcal{AB}}$$

Since the emitter is moving, the latter signal needs to travel for a longer distance to be received,

$$t_{\mathcal{CD}} = t_{\mathcal{AB}} + \frac{v}{c} t_{\mathcal{AB}} = \gamma\left(1 + \frac{v}{c}\right) t'_{\mathcal{AB}} = \sqrt{\frac{1+\beta}{1-\beta}}\, t'_{\mathcal{AB}} > t'_{\mathcal{AB}}$$

Therefore we have $\nu < \nu'$, a redshift. If the emitter moves towards the observer, we have a blueshift.

1.8 Velocity Addition

Velocity addition arises naturally from the Lorentz Transformation. For some trajectory $x^i(t)$, we just want to find how $\frac{\mathrm{d}x'^i}{\mathrm{d}t'}$ and $\frac{\mathrm{d}x^i}{\mathrm{d}t}$ are related.

Starting from the differentials of coordinates for two inertial frames S and S' with relative speed v in x-direction:

$$\begin{aligned} \mathrm{d}t' &= \gamma_v \left(\mathrm{d}t - v\mathrm{d}x/c^2\right) \\ \mathrm{d}x' &= \gamma_v(\mathrm{d}x - v\mathrm{d}t) \\ \mathrm{d}y' &= \mathrm{d}y \\ \mathrm{d}z' &= \mathrm{d}z \end{aligned}$$

we easily have

$$\begin{aligned} u'_x &= \frac{\mathrm{d}x'}{\mathrm{d}t'} = \frac{u_x - v}{1 - u_x v/c^2} \\ u'_y &= \frac{\mathrm{d}y'}{\mathrm{d}t'} = \frac{u_y}{\gamma_v\left(1 - u_x v/c^2\right)} \\ u'_z &= \frac{\mathrm{d}z'}{\mathrm{d}t'} = \frac{u_z}{\gamma_v\left(1 - u_x v/c^2\right)} \end{aligned} \tag{11}$$

The inverse can be easily found similarly.

From above, we can see that the composition of velocity is non-linear in Special Relativity as there is a uniform upper limit of speed for moving object, c. However, for addition of velocities for co-linear motions, we can see that the transformation is linear in rapidity since this can be seen as composition of hyperbolic "rotations". For co-linear v, u', and composite u, $\psi_u = \psi_v + \psi_{u'}$, we have

$$u = c\tanh(\psi_v + \psi_{u'}) = c\frac{\tanh\psi_v + \tanh\psi_{u'}}{1 + \tanh\psi_v \tanh\psi_{u'}} = \frac{u' + v}{1 + u'v/c^2}$$

1.9 Acceleration in Special Relativity

Till now, we kept dealing with constant velocities. Although the postulates of Special Relativity are built on the notion of inertial frames, we can still study acceleration in this context.

1.9.1 Acceleration appears to observers

For different observers in inertial frames S and S' with relative motion of constant speed v in x-direction, the trajectory of some particle P is denoted $x^i(t)$ and $x'^i(t')$, respectively. Then the observers measure the acceleration of the particle as

$$a^i = \frac{\mathrm{d}u^i}{\mathrm{d}t} = \frac{\mathrm{d}^2 x^i}{\mathrm{d}t^2} \qquad \text{and} \qquad a'^i = \frac{\mathrm{d}u'^i}{\mathrm{d}t'} = \frac{\mathrm{d}^2 x'^i}{\mathrm{d}t'^2}$$

respectively.

Considering the differentials of velocity transformations:

$$\begin{aligned} \mathrm{d}u'_x &= \frac{\mathrm{d}u_x}{\gamma_v^2\left(1 - u_x v/c^2\right)^2} \\ \mathrm{d}u'_y &= \frac{\mathrm{d}u_y}{\gamma_v\left(1 - u_x v/c^2\right)} + \frac{u_y v \mathrm{d}u_x}{c^2\gamma_v\left(1 - u_x v/c^2\right)^2} \\ \mathrm{d}u'_z &= \frac{\mathrm{d}u_z}{\gamma_v\left(1 - u_x v/c^2\right)} + \frac{u_z v \mathrm{d}u_x}{c^2\gamma_v\left(1 - u_x v/c^2\right)^2} \end{aligned}$$

after some tedious calculations, we get

$$\begin{aligned} a'_x &= \frac{\mathrm{d}u'_x}{\mathrm{d}t'} = \frac{1}{\gamma_v^3\left(1 - u_x v/c^2\right)^3} a_x \\ a'_y &= \frac{\mathrm{d}u'_y}{\mathrm{d}t'} = \frac{1}{\gamma_v^2\left(1 - u_x v/c^2\right)^2} a_y + \frac{u_y v}{c^2\gamma_v^2\left(1 - u_x v/c^2\right)^3} a_x \\ a'_z &= \frac{\mathrm{d}u'_z}{\mathrm{d}t'} = \frac{1}{\gamma_v^2\left(1 - u_x v/c^2\right)^2} a_z + \frac{u_z v}{c^2\gamma_v^2\left(1 - u_x v/c^2\right)^3} a_x \end{aligned} \tag{12}$$

Although the acceleration transforms in such a complicated way, it is an absolute notion for any observer since they would all agree on if any particle is accelerating.

1.9.2 Acceleration appears to a particle

> DEFINITION 1.13 (Instantaneous rest frame). An instantaneous rest frame for some particle P, at each instant, is an inertial frame co-moving with P.

If we denote S' as the instantaneous rest frame for some accelerating particle, then at each instant, $u' = 0, t' = \tau$. For the particle itself, its own "accelerometer" reads an acceleration $f(\tau)$ at some proper time τ. For S', this is simply

$$\frac{\mathrm{d}u'}{\mathrm{d}\tau} = f(\tau)$$

For some observer at some rest frame S, it measures

$$\frac{\mathrm{d}u}{\mathrm{d}t} = \frac{1}{\gamma_u^3} f(\tau) \qquad \Rightarrow \qquad \frac{\mathrm{d}u}{\mathrm{d}\tau} = \frac{1}{\gamma_u^2} f(\tau)$$

The acceleration marked by S is indeed non-linear. But with the notion of rapidity, we can see

$$\boxed{c\frac{\mathrm{d}\psi}{\mathrm{d}\tau} = f(\tau)} \tag{13}$$

where it gives a relativistic analogous of Newtonian velocity-acceleration relationship.

We find

$$c\psi(\tau) = \int_0^\tau f(\tau')\,\mathrm{d}\tau' + c\psi_0$$

with ψ_0 some initial rapidity.

Also, it is feasible to mark the trajectory of the accelerating particle (later, known as the *worldline*) using τ as the parameter, since

$$\frac{\mathrm{d}t}{\mathrm{d}\tau} = \cosh\psi(\tau) \qquad \text{and} \qquad \frac{\mathrm{d}x}{\mathrm{d}\tau} = c\sinh\psi(\tau)$$

where we integrate to get $t(\tau)$ and $x(\tau)$.

EXAMPLE. For acceleration f constant, in 2D spacetime, we have the trajectory of the particle (started from spacetime origin) as

$$t = \frac{c}{f}\sinh\frac{f\tau}{c}$$
$$x = \frac{c^2}{f}\left(\cosh\frac{f\tau}{c} - 1\right)$$

appeared to be a hyperbola. It has an asymptotic line $ct = c^2/f + x$, marking the *event horizon* for this accelerating particle, since by the causal structure of light cone, the events beyond the event horizon cannot causally communicate with or influence the particle.

2 Manifolds and Tensors

In the following three sections, I will introduce some mathematics using which we can better describe the physics in General Relativity. Please note that many of my constructions are not rigorous enough, however, these would well satisfy the purpose of this course. Also, I may become "sloppy" in notations, but these are, under most circumstances, for the clearer illustration of intuitions.

The first two classes of objects to be introduced are manifolds and tensors. The whole motivation of studying manifolds might be unclear until we introduce the equivalence principle that leads to the formulation of GR, later in this course. Please bear in mind now. However, after defining tensors, we can immediately see why they are good mathematical objects which we can use to describe physical quantities.

2.1 Manifolds

DEFINITION 2.1 (Manifold). An n-dimensional manifold (n-manifold) is a set $\mathcal{M}$ that locally resembles $\mathbb{R}^n$, i.e. for any $p \in \mathcal{U} \subset \mathcal{M}$, where $\mathcal{U}$ is open, there exists a bijective and bi-continuous map

$$\phi : \mathcal{U} \to D \subset \mathbb{R}^n, p \mapsto (x^1, x^2, \cdots, x^n)$$

where D is open. ϕ is called a chart. $\{x^\alpha\}, \alpha = 1, 2, \cdots, n$ are called coordinates in the chart.

NOTE. To complete the above definition more rigorously, we require

(i) Manifolds are well-behaved topological spaces.

(ii) Charts are compatible, i.e. transition functions $\phi_\beta \circ \phi_\alpha^{-1}$ are smooth for any charts ϕ_α, ϕ_β.

(iii) There exists a set of compatible charts that covers $\mathcal{M}$, called an atlas, i.e. domains $\mathcal{U}_\alpha$ of such charts ϕ_α satisfy

$$\bigcup_\alpha \mathcal{U}_\alpha = \mathcal{M}$$

With this definition, we can always map objects on the manifold to certain correspondences on $\mathbb{R}^n$ and use our existing knowledge valid on $\mathbb{R}^n$ to tackle problems.

EXAMPLES.

(i) Spherical polar coordinates provide a chart on a unit sphere. It cannot cover the whole sphere since there are two singularities at $\theta = 0$ and $\theta = \pi$.

(ii) Passive coordinate transformation $x \to x'$ with $x'^\alpha(x^\beta), \alpha, \beta = 1, \cdots, n$. This shows that we may cover the same open subset of a manifold with different charts. We note that the differentials of coordinates transforms like

$$\mathrm{d}x'^\alpha = \frac{\partial x'^\alpha}{\partial x^\beta} \mathrm{d}x^\beta$$

where we note the Jacobian is

$$J = \det\left(\frac{\partial x'^{\alpha}}{\partial x^{\beta}}\right)$$

We can just use our knowledge from calculus on $\mathbb{R}^n$.

We can think that an n-manifold is embedded in high-dimensional Euclidean space, but we should appreciate the existence of such a manifold independent of embedding. (There is a theorem shows that any n-manifold can be shown to be embedded in $\mathbb{R}^{2n}$.)

2.2 Functions and Curves

Definition 2.2 (Function). A function f on a manifold $\mathcal{M}$ is a map $f : \mathcal{M} \to \mathbb{R}$. f is smooth iff for any chart ϕ, $f \circ \phi^{-1} : D \subset \mathbb{R}^n \to \mathbb{R}$ is smooth. If f is independent of coordinate transformation, we call it a scalar.

Definition 2.3 (Curve). A curve on manifold $\mathcal{M}$ is a map $\lambda : I \subset \mathbb{R} \to \mathcal{M}$, I is open. λ is smooth iff for any relevant chart ϕ, $\phi \circ \lambda : I \to D \subset \mathbb{R}^n$ is smooth.

Definition 2.4 (Space of smooth functions). We denote $C^\infty(\mathcal{M})$ the space of all smooth functions $f : \mathcal{M} \to \mathbb{R}$.

2.3 Vectors and Vector Fields

Definition 2.5 (Vector). A (tangent) vector X_p at point $p \in \mathcal{M}$ is a map

$$X_p : C^\infty(\mathcal{M}) \to \mathbb{R}$$

satisfying:

(i) Linearity: $X_p(f+g) = X_p(f) + X_p(g), \forall f, g \in C^\infty(\mathcal{M})$;

(ii) $X_p(f) = 0$ when f is constant;

(iii) Leibniz rule: $X_p(fg) = X_p(f)g(p) + f(p)X_p(g), \forall f, g \in C^\infty(\mathcal{M})$.

Definition 2.6 (Tangent space). The tangent space of manifold $\mathcal{M}$ at point $p \in \mathcal{M}$ is the space of all vectors at p. Denoted $T_p(\mathcal{M})$. $T_p(\mathcal{M})$ forms a vector space over $\mathbb{R}$.

Note. From the above definition of vectors, we see that they are equivalent to differential operators. For each X_p at $p \in \mathcal{M}$, we can always find a smooth curve λ passes p with $\lambda(0) = p$, such that

$$X_p : f \mapsto \frac{\mathrm{d}}{\mathrm{d}t} f(\lambda(t))\bigg|_{t=0}$$

This is actually identical to the idea of a local vector (e.g. tangent vector to a curve or electric field) we've met before. A vector, intuitively, should have magnitude and direction. Vectors can be represented by differential operators $\frac{\mathrm{d}}{\mathrm{d}t}$ since, informally speaking, we can choose $\lambda(t)$ such that it is tangent to the direction of the desired vector and we can tune the scale density in t to present the magnitude.

However, there is another class of vectors, known as position vectors. These vectors are only valid in Euclidean spaces, since the set of position vectors is isomorphic to the global coordinate system. For a general manifold, we cannot find a single global chart to represent each point, thus position vectors are no longer valid.

As now all vectors on a manifold are local vectors, for different points, the vectors live in different mathematical spaces. There is no mathematical or physical meaning to add two vectors from different points.

For a chart ϕ with coordinates $\{x^\alpha\}$, for some $f \in C^\infty(\mathcal{M})$, we can write

$$X_p(f) = \frac{\mathrm{d}}{\mathrm{d}t} f(\lambda(t))\bigg|_{t=0} = \frac{\mathrm{d}x^\alpha}{\mathrm{d}t}\frac{\partial f}{\partial x^\alpha}\bigg|_{t=0}$$

for some valid λ. This is an instance of chain rule

$$\text{“}\frac{\mathrm{d}}{\mathrm{d}t} = \frac{\mathrm{d}x^\alpha}{\mathrm{d}t}\frac{\partial}{\partial x^\alpha}\text{”}$$

leading us to think if $\{\frac{\partial}{\partial x^\alpha}\}$ form a basis of the tangent space.

THEOREM 2.1. *For an n-manifold $\mathcal{M}$, at $p \in \mathcal{M}$, under coordinates $\{x^\alpha\}$*

$$\partial_\alpha\bigg|_p := \frac{\partial}{\partial x^\alpha}\bigg|_p$$

provides a basis for $T_p(\mathcal{M})$, called the coordinate basis. This means $\dim T_p(\mathcal{M}) = n$.

Here I won't show the proof as it is beyond our concern. It can be found from [3].

From now on, I will omit the evaluation symbols to save some handwriting.

With the above theorem, under certain coordinate basis $\{\partial_\alpha\}$, we can write

$$X_p = X_p^\alpha \partial_\alpha$$

where $X_p^\alpha = \frac{\mathrm{d}x^\alpha}{\mathrm{d}t}$ are identified as components of X_p under that coordinate basis.

We now explore the behaviour of vectors under coordinate transformation $x^\alpha \to x'^\alpha$. By chain rule, the basis transform as

$$\frac{\partial}{\partial x'^\alpha} = \frac{\partial x^\beta}{\partial x'^\alpha}\frac{\partial}{\partial x^\beta} \tag{14}$$

The components transform as

$$X_p'^\alpha = \frac{\mathrm{d}x'^\alpha}{\mathrm{d}t} = \frac{\partial x'^\alpha}{\partial x^\beta}\frac{\mathrm{d}x^\beta}{\mathrm{d}t} = \frac{\partial x'^\alpha}{\partial x^\beta}X_p^\beta \tag{15}$$

Overall, it's easy to see

$$X_p' = X_p'^\alpha \partial_\alpha' = X_p^\alpha \partial_\alpha = X_p$$

showing that vectors are coordinate independent. This is also obvious from the fact that X_p only depends on the curve $\lambda(t)$, which is itself independent of coordinates.

The independence of coordinates allows us to use vectors on manifolds to represent certain physical quantities. This is a nice structure, naturally compatible with the principle of relativity, as we will see later in the course.

Definition 2.7 (Vector field). A vector field X is a map $X : C^\infty(\mathcal{M}) \to C^\infty(\mathcal{M})$, with

$$X(f)(p) = X_p(f)$$

We denote $V(\mathcal{M})$ the space of all vector fields on $\mathcal{M}$.

Later in the course, we may tend to call vector fields simply "vectors". Although this may be ambiguous, under most circumstances it should be clear from context. Also, for vectors, we would omit the subscript p as it reduces the clarity when we use indices notations.

Definition 2.8 (Integral curve). An integral curve $\lambda(t)$ of a vector field X through a point $p \in \mathcal{M}$ is the curve through p whose tangent vector at every point q along the curve is X_q.

We know that $\frac{\mathrm{d}}{\mathrm{d}t}\big|_\lambda = X$, in coordinate basis, the equation

$$\frac{\mathrm{d}x^\alpha(\lambda(t))}{\mathrm{d}t} = X^\alpha(\lambda(t))$$

with $x^\alpha(\lambda(t_0)) = x^\alpha(p)$ has a unique solution by ODE theory.

2.4 Covectors and 1-Forms

Definition 2.9 (Covector). A cotangent vector, or in short as a covector, η at point $p \in \mathcal{M}$ is a map

$$\eta : T_p(\mathcal{M}) \to \mathbb{R}$$

satisfying:

(i) Linearity: $\eta(aX + bY) = a\eta(X) + b\eta(Y)$, $\forall X, Y \in T_p(\mathcal{M}), a, b \in \mathbb{R}$;

(ii) Scalar property: $\eta(X)$ is coordinate independent, $\forall X \in T_p(M)$.

We define $\eta_\mu = \eta(\partial_\mu)$ the component of η under coordinate basis.

Definition 2.10 (Cotangent space). For manifold $\mathcal{M}$, the cotangent space $T_p^*(\mathcal{M})$ at point $p \in \mathcal{M}$ is the space of all covectors at p. It is the dual space of $T_p(\mathcal{M})$, also a vector space over $\mathbb{R}$.

To see how we can construct a coordinate basis for $T_p^*(\mathcal{M})$, consider the following definition.

Definition 2.11 (Gradient). The gradient of a function $f \in C^\infty(\mathcal{M})$ is a map

$$\mathrm{d}f : T_p(\mathcal{M}) \to \mathbb{R}, \quad X \mapsto X(f)$$

By the linearity of vectors and the coordinate independence, it is easy to see $\mathrm{d}f \in T_p^*(\mathcal{M})$. Coordinates x^α can also be seen as functions and we can take their gradient $\mathrm{d}x^\alpha$ to see

$$\mathrm{d}x^\alpha(\partial_\beta) = \frac{\partial x^\alpha}{\partial x^\beta} = \delta^\alpha{}_\beta$$

where ${\delta^\alpha}_\beta$ is the Kronecker delta that

$${\delta^\alpha}_\beta = \begin{cases} 1, & \alpha = \beta \\ 0, & \alpha \neq \beta \end{cases}$$

This means

$$\eta(\partial_\alpha) = \eta_\alpha = \eta_\beta {\delta^\beta}_\alpha = \eta_\beta \mathrm{d}x^\beta(\partial_\alpha)$$

i.e.

$$\eta = \eta_\alpha \mathrm{d}x^\alpha$$

making $\{\mathrm{d}x^\alpha\}$ the coordinate basis of $T_p^*(\mathcal{M})$. Also it is easy to verify $\eta(X) = \eta_\alpha X^\alpha$

Since at $p \in \mathcal{M}$, $\forall X \in T_p(\mathcal{M}), \eta \in T_p^*(\mathcal{M})$, $\eta(X)$ should be coordinate independent, i.e. for coordinate transformation $x^\alpha \to x'^\alpha$,

$$\eta'_\alpha X'^\alpha = \eta'_\alpha \frac{\partial x'^\alpha}{\partial x^\beta} X^\beta = \eta_\beta X^\beta$$

This means the components of covectors transform as

$$\eta'_\alpha = \frac{\partial x^\beta}{\partial x'_\alpha} \eta_\beta \tag{16}$$

It is also clear that the coordinate basis $\{\mathrm{d}x^\alpha\}$ transform exactly like the differentials of coordinates we met before.

Definition 2.12 (1-form). A 1-form or a covector field ω on manifold $\mathcal{M}$ is a map $\omega : V(\mathcal{M}) \to \mathbb{R}$. We denote the space of all 1-forms on $\mathcal{M}$ as $\Lambda^1(\mathcal{M})$.

1-form is an instance of a class of objects called differential forms. They are useful mathematical ideas for physics. But in this course we will not discuss them.

Above, we've mentioned that $T_p^*(\mathcal{M})$ is a dual space of $T_p(\mathcal{M})$. By the definition of a dual space, $T_p^*(\mathcal{M})$ is the space of all linear maps $T_p(\mathcal{M}) \to \mathbb{R}$. By Linear Algebra, it can be shown that $T_p(\mathcal{M})$ and $T_p^*(\mathcal{M})$ are isomorphic. Further more, it is obvious that the dual of $T_p^*(M)$, denoted $T_p^{**}(\mathcal{M})$, is isomorphic to $T_p(\mathcal{M})$.

In next section, we will introduce a new object called metric, which establishes a natural isomorphism between $T_p(\mathcal{M})$ and $T_p^*(\mathcal{M})$.

2.5 Tensors

Definition 2.13 (Tensor). A tensor T at $p \in \mathcal{M}$ of rank (r, s), $r, s \in \mathbb{N}_0$, is a multilinear map

$$T : \underbrace{T_p^*(\mathcal{M}) \times \cdots \times T_p^*(\mathcal{M})}_{r \text{ copies}} \times \underbrace{T_p(\mathcal{M}) \times \cdots \times T_p(\mathcal{M})}_{s \text{ copies}} \to \mathbb{R}$$

It is said to have total rank $r + s$. We define the components of T as

$$T^{\alpha_1 \cdots \alpha_r}{}_{\beta_1 \cdots \beta_s} = T(\mathrm{d}x^{\alpha_1}, \cdots, \mathrm{d}x^{\alpha_r}; \partial_{\beta_1}, \cdots, \partial_{\beta_s})$$

under coordinate basis.

Examples.

(i) Any covector η is a tensor of rank $(0,1)$.

(ii) Any vector X is a tensor of rank $(1,0)$.

(iii) For a tensor $\delta : T_p^*(\mathcal{M}) \times T_p(\mathcal{M}) \to \mathbb{R}$, $(\eta, X) \mapsto \eta(X)$, $\forall \eta \in T_p^*(\mathcal{M})$, $X \in T_p(\mathcal{M})$ is a tensor of rank $(1,1)$ with components $\delta^\alpha{}_\beta = \delta(\mathrm{d}x^\alpha, \partial_\beta) = \partial x^\alpha / \partial x^\beta =$ Kronecker Delta.

It can be shown that the components of a (r,s)-tensor transform under $x^\alpha \to x'^\alpha$ as

$$T'^{\alpha_1 \cdots \alpha_r}{}_{\beta_1 \cdots \beta_s} = \frac{\partial x'^{\alpha_1}}{\partial x^{\mu_1}} \cdots \frac{\partial x'^{\alpha_r}}{\partial x^{\mu_r}} \frac{\partial x^{\nu_1}}{\partial x'^{\beta_1}} \cdots \frac{\partial x^{\nu_s}}{\partial x'^{\beta_s}} T^{\mu_1 \cdots \mu_r}{}_{\nu_1 \cdots \nu_s} \tag{17}$$

Definition 2.14 (Tensor field). A tensor field of rank (r,s) is a smooth assignment of (r,s)-tensors at each point $p \in \mathcal{M}$, i.e. $T : p \mapsto T_p$. The tensor field is smooth iff its components in any coordinate basis are smooth functions.

2.6 Tensor Operations

Practically, it is beneficial to know how we can play around with tensors properly and efficiently as they are in general very complicated objects.

Tensors have the following operations:

(1) Addition and scalar multiplication: For example, let $c_1, c_2 \in \mathbb{R}$, S, T are $(1,1)$-tensors, we have

$$c_1 S + c_2 T : T_p^*(\mathcal{M}) \times T_p(\mathcal{M}) \to \mathbb{R}, \quad (\eta, X) \mapsto c_1 S(\eta, X) + c_2 T(\eta, X).$$

(2) (Anti-)Symmetrization: (in examples)

  (i) Symmetric part $S_{\alpha\beta} := \frac{1}{2}(T_{\alpha\beta} + T_{\beta\alpha}) =: T_{(\alpha\beta)}$

  (ii) Anti-symmetric part $A_{\alpha\beta} := \frac{1}{2}(T_{\alpha\beta} - T_{\beta\alpha}) =: T_{[\alpha\beta]}$

  (iii) Index subset: e.g. $T^{(\alpha\beta)\gamma}{}_\delta := \frac{1}{2}(T^{\alpha\beta\gamma}{}_\delta + T^{\beta\alpha\gamma}{}_\delta)$

  (iv) Non-adjacent indices: e.g. $T_{(\alpha|\beta\gamma|\delta)} := \frac{1}{2}(T_{\alpha\beta\gamma\delta} + T_{\delta\beta\gamma\alpha})$

  (v) Over $n > 2$ indices, sum over all permutations. Sign of permutation should be applied for anti-symmetrization. A factor of $n!$ should be divided by. E.g.

$$T^\alpha{}_{[\beta\gamma\delta]} = \frac{1}{3!}(T^\alpha{}_{\beta\gamma\delta} + T^\alpha{}_{\gamma\delta\beta} + T^\alpha{}_{\delta\beta\gamma} - T^\alpha{}_{\gamma\beta\delta} - T^\alpha{}_{\delta\gamma\beta} - T^\alpha{}_{\beta\delta\gamma})$$

(3) Contraction of (r,s)-tensor results in a $(r-1, s-1)$-tensor: sum over one upstairs and one downstairs index.

Example. Let T be a $(3,2)$-tensor. It can give a $(2,1)$-tensor S by defining

$$S(\omega, \eta, X) := T(\mathrm{d}x^\mu, \omega, \eta, \partial_\mu, X) \quad \text{(sum over } \mu\text{)}$$

This is basis independent since $\mathrm{d}x^\alpha(\partial_\beta) = \delta^\alpha{}_\beta$, $\mathrm{d}x'^\mu(\partial'_\nu) = \delta^\mu{}_\nu$ and $\mathrm{d}x'^\mu = \frac{\partial x'^\mu}{\partial x^\alpha}\mathrm{d}x^\alpha$, we have

$$T(\mathrm{d}x'^\mu, \omega, \eta, \partial'_\mu, X) = \underbrace{\frac{\partial x'^\mu}{\partial x^\alpha}\frac{\partial x^\beta}{\partial x'^\mu}}_{\delta^\beta{}_\alpha} T(\mathrm{d}x^\alpha, \omega, \eta, \partial_\beta, X)$$

is a tensor. The components are

$$S^{\mu\nu}{}_\rho = T^{\alpha\mu\nu}{}_{\alpha\rho}$$

(4) Tensor product: Let S be a (p, q)-tensor and T a (r, s)-tensor. The tensor product $S \otimes T$ is a $(p + r, q + s)$-tensor.

$$(S \otimes T)(\omega_1, \cdots, \omega_p, \eta_1, \cdots, \eta_r, X_1, \cdots, X_q, Y_1, \cdots, Y_s) := \\ S(\omega_1, \cdots, \omega_p, X_1, \cdots, X_q)T(\eta_1, \cdots, \eta_r, Y_1, \cdots, Y_s)$$

One shows

(i) $(S \otimes T)^{\alpha_1\cdots\alpha_p\beta_1\cdots\beta_r}{}_{\mu_1\cdots\mu_q\nu_1\cdots\nu_s} = S^{\alpha_1\cdots\alpha_p}{}_{\mu_1\cdots\mu_q} T^{\beta_1\cdots\beta_r}{}_{\nu_1\cdots\nu_s}$

(ii) E.g. in a coordinate basis, a $(2, 1)$-tensor T can be written

$$T = T^{\mu\nu}{}_\rho\,(\partial_\mu \otimes \partial_\nu \otimes \mathrm{d}x^\rho)$$

Likewise for (r, s)-tensor.

As shown by the tensor product operation above, any general tensor can be constructed by a tensor product of vectors and covectors. Since vectors and covectors are coordinate independent, tensors are essentially coordinate independent, too. They provide suitable mathematical objects using which we can describe certain physical quantities. As the perspectives (bases) and observations (components) change, the underlying physics (tensors) remains the same for all observers.

In next section, we will introduce two new mathematical objects: metric and connection. I will show how we can do calculus on manifolds with them.

3 Metric and Connection

In the previous section, we developed the concept of manifolds and tensors. By seeing the generality of manifolds and the coordinates-independent nature of tensors, we consider them as good candidates in describing the physics of GR later in the course. However, they are not enough for the whole story. We need to construct two more important things on manifolds: the metric and the connection. The former gives us a measure of distance, the latter tells us how tensor fields change throughout a manifold.

3.1 Metric Tensor

After introducing the manifold, we would like to know how far any two points on the manifold are separated. Recall our definition of the spacetime interval in section 1, we hereby can construct a more general version of it by assuming the infinitesimal distance squared is related to differentials in coordinates quadratically:

$$\mathrm{d}s^2 = g_{\mu\nu}(x)\mathrm{d}x^\mu \mathrm{d}x^\nu \tag{18}$$

By this construction, we immediately know that the coefficients $g_{\mu\nu}$ should be symmetric in its two indices. By requiring $\mathrm{d}s^2$ invariant under coordinate transformation, the coefficients $g_{\mu\nu}$ transform as components of a $(0, 2)$-tensor. This is left as an exercise to check.

Thus, we have the following formal definition.

Definition 3.1 (Metric). A metric on manifold $\mathcal{M}$ is a $(0, 2)$-tensor field g that is

(i) Symmetric: $g(X, Y) = g(Y, X), \forall X, Y \in V(\mathcal{M})$;

(ii) Non-degenerate: for some $p \in \mathcal{M}$, $g(X, Y)\big|_p = 0$ for $\forall Y \in T_p(\mathcal{M})$ iff $X_p = 0$.

Under coordinate basis, we have

$$g = g_{\mu\nu}\mathrm{d}x^\mu \otimes \mathrm{d}x^\nu$$

Since $g_{\mu\nu}$ is symmetric, when viewed as a matrix, we can diagonalise it under certain orthogonal basis. Since g is non-degenerate, we note that after diagonalisation, all its diagonal entries are non-zero. At some point $p \in \mathcal{M}$, we can diagonalise and rescale $g_{\mu\nu}$ such that

$$g_{\mu\nu}\bigg|_p = \mathrm{diag}(\pm 1, \pm 1, \cdots, \pm 1)$$

Here we have a theorem regarding this.

Theorem 3.1 (Sylvester's law of inertia). *The number of positive and negative elements in a diagonalised matrix is independent of the choice of orthonormal basis.*

This motivates the following definitions.

DEFINITION 3.2 (Signature). The signature is the number of positive and negative elements of the metric.

DEFINITION 3.3 (Riemannian/Lorentzian metric). A metric is Riemannian if the signature is $(++\cdots+)$. A metric is Lorentzian if the signature is $(+-\cdots-)$.

NOTE. The signature $(+-\cdots-)$ for Lorentzian metric is just a convention we adopted. In the other convention, it can be $(-+\cdots+)$, which we shall not use in this course.

DEFINITION 3.4 (Riemannian/Lorentzian manifold). A Riemannian/Lorentzian manifold is a pair $(\mathcal{M}, g)$ where $\mathcal{M}$ is a manifold and g is a Riemannian/Lorentzian metric. A Lorentzian manifold can also be called a spacetime.

EXAMPLES.

(i) A 3D Euclidean space $\mathbb{R}^3$ is a Riemannian manifold with metric

$$g = \mathrm{d}x^1 \otimes \mathrm{d}x^1 + \mathrm{d}x^2 \otimes \mathrm{d}x^2 + \mathrm{d}x^3 \otimes \mathrm{d}x^3$$

(ii) A 4D Minkowski spacetime is a Lorentzian manifold with metric

$$\eta = \mathrm{d}x^0 \otimes \mathrm{d}x^0 - \mathrm{d}x^1 \otimes \mathrm{d}x^1 - \mathrm{d}x^2 \otimes \mathrm{d}x^2 - \mathrm{d}x^3 \otimes \mathrm{d}x^3$$

Since the metric tensor is non-degenerate, the components viewed as a matrix is invertible, motivating our definition of an inverse metric.

DEFINITION 3.5 (Inverse metric). For a metric g on manifold $\mathcal{M}$, its inverse metric is a $(2,0)$-tensor field $\hat{g}$, expressed in coordinate basis as

$$\hat{g} := g^{\mu\nu} \partial_\mu \otimes \partial_\nu$$

such that its components satisfy

$$g^{\mu\nu} g_{\nu\rho} = \delta^\mu{}_\rho$$

3.2 Benefits of Having a Metric

After constructing metric, we have more interesting things on a manifold.

3.2.1 Natural Isomorphism between Vectors and Covectors

Indeed, vectors and covectors on a manifold are distinct mathematical objects. How they are related is vague before the construction of a metric. Now, using the metric, we can establish a natural isomorphism between vectors and covectors.

Use g, we establish

$$g_\downarrow : T_p(\mathcal{M}) \to T_p^*(\mathcal{M}), \quad X \mapsto g(X, \cdot)$$

where "$\cdot$" is just some place holder. Or in components

$$X_\mu = g_{\mu\nu} X^\nu \tag{19}$$

Similarly, with the inverse metric, we establish

$$\hat{g}_{\uparrow} : T_p^*(\mathcal{M}) \to T_p(\mathcal{M}), \quad \eta \mapsto \hat{g}(\eta, \cdot)$$

or in components

$$\eta^\mu = g^{\mu\nu}\eta_\nu \tag{20}$$

It is easy to check the pair $g_{\downarrow}, \hat{g}_{\uparrow}$ gives an isomorphism between vectors and covectors at any point on a manifold. (Exercise: check this!) Physicists often call these operations "raising/lowering the indices".

3.2.2 Scalar Product

With a metric, we can generalise the notion of scalar product of vectors from Euclidean space to a general manifold (for vectors at the same point).

> Definition 3.6 (Scalar product). On a Riemannian/Lorentzian manifold $(\mathcal{M}, g)$, the scalar product of two vectors X, Y at $p \in \mathcal{M}$ is
>
> $$\langle X, Y \rangle := g(X, Y)$$

We often write it as $\langle X, Y \rangle = g_{\mu\nu}X^\mu Y^\nu = X_\mu Y^\mu = X^\mu Y_\mu$.

Note that for a Riemannian manifold, the scalar product of any two vectors at the same point is non-negative. However, this is not the case for Lorentzian manifold. The difference in sign motivates us to define different names for vectors with distinct properties.

> Definition 3.7 (Properties of vectors on Lorentzian manifold). For a Lorentzian manifold $(\mathcal{M}, g)$, at some point $p \in \mathcal{M}$, the vector X is said to be
>
> (i) timelike, when $g(X, X) > 0$;
>
> (ii) null or lightlike, when $g(X, X) = 0$;
>
> (iii) spacelike, when $g(X, X) < 0$.

Recall the properties (timelike, null, spacelike) of spacetime interval $\mathrm{d}s^2$ back in section 1. The same definition for an interval on a Lorentzian manifold can be drawn from the properties of vectors tangent to it.

When we define the norm of a vector, we need to consider the absolute value of the scalar product of the vector with itself.

> Definition 3.8 (Norm of vector). The norm of a vector X is
>
> $$|X| := |g(X, X)|^{1/2}$$

Also we can extend the idea of the angle between conventional vectors.

> Definition 3.9 (Angle between vectors). The angle θ between two non-null vectors X, Y is defined by
>
> $$\cos\theta := \frac{g(X, Y)}{|X||Y|}$$

Note that $|\cos\theta| > 1$ is possible for vectors on a Lorentzian manifold.

3.2.3 Length and Volume

The metric gives the measure of distance between two infinitesimally close points. The intuition is to integrate them to get the length of a finite segment of a curve. Use appropriate parameterisation t, the length of curve $\lambda(t)$, with tangent vector $X = \mathrm{d}/\mathrm{d}t$, between $t = t_A$ and $t = t_B$ can be calculated by

$$L_{AB} = \int_{t_A}^{t_B} \sqrt{|g(X,X)|\Big|_{\lambda(t)}}\,\mathrm{d}t = \int_{t_A}^{t_B} \left|g_{\mu\nu}\frac{\mathrm{d}x^\mu}{\mathrm{d}t}\frac{\mathrm{d}x^\nu}{\mathrm{d}t}\right|^{1/2}\mathrm{d}t \tag{21}$$

in coordinates $\{x^\mu\}$.

Note that in the above operation, all information of the manifold is translated by the appropriately chosen charts to Euclidean, where we can use multivariable calculus. It would be clear from the context when we work in $\mathbb{R}^n$. (The ultimate purpose of introducing Differential Geometry is that they enable us to eventually work in $\mathbb{R}^n$ in a rigorous manner when studying general manifolds.)

Similarly, using the metric, we can study the volume element of a certain region on a manifold. Now consider the general volume element in an n-manifold with metric g.

As $g_{\mu\nu}$ is symmetric, using appropriate coordinates $\{x^\mu\}$, we can diagonalise it such that $g_{\mu\nu} = \mathrm{diag}(g_{11}, g_{22}, \cdots, g_{nn})$ and

$$\mathrm{d}s^2 = g_{11}(\mathrm{d}x^1)^2 + g_{22}(\mathrm{d}x^2)^2 + \cdots + g_{nn}(\mathrm{d}x^n)^2$$

In this case, the basis are orthogonal (by the properties of diagonalised matrices). For each direction, the length element is

$$\mathrm{d}s^\alpha = \sqrt{|g_{\alpha\alpha}|}\mathrm{d}x^\alpha \qquad \text{(for some fixed } \alpha\text{)}$$

Thus, by orthogonality of the basis, the volume element is

$$\mathrm{d}V = \sqrt{|g_{11}g_{22}\cdots g_{nn}|}\mathrm{d}x^1\mathrm{d}x^2\cdots\mathrm{d}x^n$$

Write in a coordinate independent way, we have

$$\mathrm{d}V = \sqrt{|g|}\mathrm{d}x^1\mathrm{d}x^2\cdots\mathrm{d}x^n \tag{22}$$

where $g := \det(g_{\mu\nu})$, when diagonalised, $\det(g_{\mu\nu}) = g_{11}g_{22}\cdots g_{nn}$. Here we use the same notation as the tensor g, but which we refer to would be clear from the context.

Also, by constraining $x^\alpha =$ constant for certain α, we can find the volume/surface element for certain subsets of the manifold.

CLAIM 3.2. $\mathrm{d}V$ *is invariant under coordinate transformation.*

Proof. For $\mathrm{d}V = \sqrt{|g|}\mathrm{d}x^1\cdots\mathrm{d}x^n$ in coordinates $\{x^\mu\}$, consider coordinate transformation $x^\mu \to x'^\mu$, by multivariable calculus, we have

$$\mathrm{d}x'^1\cdots\mathrm{d}x'^n = |J|\mathrm{d}x^1\cdots\mathrm{d}x^n$$

where $J = \det\left(\frac{\partial x'^\mu}{\partial x^\nu}\right)$ is the Jacobian factor.

But by the tensorial nature of g, its components transform as

$$g'_{\mu\nu} = \frac{\partial x^\rho}{\partial x'^\mu}\frac{\partial x^\sigma}{\partial x'^\nu} g_{\rho\sigma}$$

By $\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B})$ for any two square matrices $\mathbf{A}, \mathbf{B}$, we have

$$g' = g/J^2$$

Then

$$\mathrm{d}V' = \sqrt{|g'|}\mathrm{d}x'^1 \cdots \mathrm{d}x'^n = \frac{\sqrt{|g|}}{|J|}|J|\mathrm{d}x^1 \cdots \mathrm{d}x^n = \mathrm{d}V$$

□

This invariance property is important. Although we cannot use a single chart to cover some general manifold, we are safe to use different coordinates in which we integrate objects with respect to the volume of the manifold. This shows that, a manifold with a metric has a unique intrinsic volume element. In more advanced Differential Geometry, the integration over manifolds is with respect to certain differential forms, and we have the freedom to choose a well-defined volume form for an orientable manifold. With a metric, we have the unique volume form we can rely on, given for free. One might choose other volume forms even if metric is given, but that's beyond our concern.

3.3 Normal Coordinates

Given a metric, it in general varies over the manifold, and may have some complicated form under general coordinates which makes calculation horrible. As we will see later in the course, it is miserable when we calculate the components of Riemann tensor through metric in general coordinates. But the good news is that a Riemannian/Lorentzian manifold, at a point, can look like an Euclidean space/Minkowski spacetime equipped with appropriate metric, under some special coordinates known as normal coordinates.

Definition 3.10 (Normal coordinates). For a Riemannian/Lorentzian manifold $(\mathcal{M}, g)$, the normal coordinates at some point $p \in \mathcal{M}$ are such that

$$g_{\mu\nu}\Big|_p = \begin{cases} \delta_{\mu\nu} = \mathrm{diag}(1, 1, \cdots, 1), & \text{Riemannian} \\ \eta_{\mu\nu} = \mathrm{diag}(1, -1, \cdots, -1), & \text{Lorentzian} \end{cases} \quad \text{and} \quad \frac{\partial g_{\mu\nu}}{\partial x^\rho}\Big|_p = 0$$

By the above definition, intuitively, we consider the metric in the neighbourhood of the chosen point in normal coordinates is a second order perturbation of Euclidean/Minkowski metric.

$$g_{\mu\nu}(x) = \begin{cases} \delta_{\mu\nu} + \mathcal{O}[(x - x_p)^2], & \text{Riemannian} \\ \eta_{\mu\nu} + \mathcal{O}[(x - x_p)^2], & \text{Lorentzian} \end{cases} \tag{23}$$

Theorem 3.3. *On a Riemannian/Lorentzian manifold $(\mathcal{M}, g)$, for $\forall p \in \mathcal{M}$, there exists normal coordinates.*

Proof. We set $n = \dim \mathcal{M}$. Consider the coordinate transformation $x^\mu \to x'^\mu$ and set $x^\mu(p) = x'^\mu(p) = 0$ for convenience. We want $\{x'^\mu\}$ to be normal coordinates.

Firstly, we require

$$g'_{\mu\nu}\bigg|_p = g_{\rho\sigma}\frac{\partial x^\rho}{\partial x'^\mu}\frac{\partial x^\sigma}{\partial x'^\nu}\bigg|_p = \delta_{\mu\nu}$$

which subjects to $\frac{1}{2}n(n+1)$ conditions as $g_{\mu\nu}$ is symmetric.

But we have n^2 degrees of freedom in choosing $\frac{\partial x^\rho}{\partial x'^\mu}\big|_p$. More degrees of freedom than conditions, we can certainly find such coordinate transformation that leads to the result of the first requirement.

For the next order, we require

$$\frac{\partial g_{\mu\nu}}{\partial x^\rho}\bigg|_p = 0$$

which subjects to $\frac{1}{2}n^2(n+1)$ conditions.

The degree of freedom in this order are determined by the term with fewest, which is proportional to $\frac{\partial^2 x^\mu}{\partial x^\nu \partial x^\rho}\Big|_p$, giving $\frac{1}{2}n^2(n+1)$ degrees of freedom. The degrees of freedom matches the conditions, we can fulfil this requirement.

There exists coordinates such that $\forall p \in \mathcal{M}$,

$$g_{\mu\nu}\bigg|_p = \delta_{\mu\nu}(\text{or } \eta_{\mu\nu}) \quad \text{and} \quad \frac{\partial g_{\mu\nu}}{\partial x^\rho}\bigg|_p = 0$$

□

Note.

(i) After enabling $g_{\mu\nu} = \delta_{\mu\nu}$ at p, there are $\frac{1}{2}n(n-1)$ degrees of freedom left. This corresponds to the dimension of orthogonal group $SO(n)$ for a Riemannian manifold or Lorentz group $SO(1, n-1)$ for a Lorentzian manifold. For a Lorentzian 4-manifold, it is 6, corresponding to the freedom of 3 rotations and 3 boosts (Lorentz transformation, recall from section 1).

(ii) For higher order terms, we don't have enough degrees of freedom to make them vanish. Especially,

$$\frac{\partial^2 g_{\mu\nu}}{\partial x^\rho \partial x^\sigma}\bigg|_p = 0$$

needs $\frac{1}{4}n^2(n+1)^2$ conditions, while $\frac{\partial^3 x^\mu}{\partial x^\nu \partial x^\rho \partial x^\sigma}$ only gives $\frac{1}{6}n^2(n+1)(n+2)$ degrees of freedom. We have

$$\frac{1}{4}n^2(n+1)^2 - \frac{1}{6}n^2(n+1)(n+2) = \frac{1}{12}n^2(n^2-1)$$

constraints unaccounted. This corresponds to the in general non-vanishing independent components of Riemann tensor, which we will meet later in the course. This means, for a Riemannian/Lorentzian manifold, in general, curvature is inevitable.

3.4 Connection

Now we consider how we can study changing tensor fields over some general manifold $\mathcal{M}$. Please note that this discussion is independent of a metric.

From the previous section, we know that tensors of the same type at different points live in different mathematical spaces, thus it is not possible to directly compare tensors at different points. However, as the components are functions, the most straightforward intuition is to check how they vary with respect to the chosen coordinates. The key is to find a **coordinate independent** object that describes the changes.

Firstly, we consider the simplest case: a function. For a function $f \in C^\infty(\mathcal{M})$ under coordinate basis $\{\partial_\mu\}$, we would like to use its partial derivatives with respect to the coordinates to study how it changes. Under coordinate transformation $x^\mu \to x'^\mu$, we have

$$\partial_\mu f \to \partial'_\mu f = \frac{\partial x^\nu}{\partial x'^\mu}\partial_\nu f$$

this transforms as the components of a $(0,1)$-tensor, which suits our requirement of coordinate independence.

Further we would like to see if partial derivatives are still good enough to describe changing vectors. Consider a vector field $X \in V(\mathcal{M})$. Under coordinate transformation $x^\mu \to x'^\mu$, we have

$$\partial_\mu X^\nu \to \partial'_\mu X'^\nu = \frac{\partial x^\rho}{\partial x'^\mu}\frac{\partial x'^\nu}{\partial x^\sigma}\partial_\rho X^\sigma + \frac{\partial x^\rho}{\partial x'^\mu}\frac{\partial^2 x'^\nu}{\partial x^\rho \partial x^\sigma}X^\sigma$$

which shows that the partial derivatives of vector components do not transform as components of a tensor. It is easy to check this is also the case for covectors.

As a tensor can be constructed by vectors and covectors, we know in general, this approach fails. This should be the fate as we were just interested how components change but we neglected the variations in basis. We want to introduce an operation that generates tensorial (i.e. coordinate independent) derivatives of tensor fields that concerns both components and basis.

The above motivation leads to the following definition.

Definition 3.11 (Connection). A connection on manifold $\mathcal{M}$ is a map

$$\nabla : V(\mathcal{M}) \times V(\mathcal{M}) \to V(\mathcal{M})$$

that satisfies

(i) $\nabla_X(Y+Z) = \nabla_X Y + \nabla_X Z$;

(ii) $\nabla_{fX+gY} Z = f\nabla_X Z + g\nabla_Y Z,\ \forall f,g \in C^\infty(\mathcal{M})$;

(iii) $\nabla_X(fY) = \nabla_X(f)Y + f\nabla_X Y,\ \forall f \in C^\infty(\mathcal{M})$ where we define $\nabla_X f = X(f)$.

by usually denoting $\nabla(X,Y)$ as $\nabla_X Y$.

Note. It is clear from the definition that ∇ is not a tensor since it is not linear in the second entry.

Definition 3.12 (Connection coefficient). Under coordinate basis $\{e_\mu = \partial_\mu\}$, the connection coefficients $\Gamma^\rho_{\mu\nu}$ of ∇ is defined by

$$\nabla_{e_\nu} e_\mu := \Gamma^\rho_{\mu\nu} e_\rho$$

For the convenience in notation, often we denote ∇_{e_ν} as ∇_ν.

Intuitively, the above construction gives a sense of how a basis vector e_μ changes along direction of e_ν. Now using ∇, we can account for changes in both components (by applying partial derivatives) and basis.

Note that the change in basis is coordinate dependent as basis vectors are so. Under coordinate transformation $x^\mu \to x'^\mu$, it can be shown the connection coefficients transform as

$$\Gamma'^{\sigma}_{\mu\nu} = \frac{\partial x'^\sigma}{\partial x^\rho}\frac{\partial x^\alpha}{\partial x'^\mu}\frac{\partial x^\beta}{\partial x'^\nu}\Gamma^{\rho}_{\alpha\beta} + \frac{\partial x'^\sigma}{\partial x^\rho}\frac{\partial^2 x^\rho}{\partial x'^\mu \partial x'^\nu} \tag{24}$$

By this transformational law, it is clear that the choice of a connection is arbitrary up to a $(1,2)$-tensor. (This also means that the difference of connections is a tensor, will be discussed later.) Later, we will see how a metric on the manifold gives us a unique choice of connection.

Now consider how vectors change using connection.

Definition 3.13 (Covariant derivative of vectors). On manifold $\mathcal{M}$ with connection ∇, the covariant derivative of $Y \in V(\mathcal{M})$ is a map

$$\nabla Y = \nabla(\cdot, Y) : V(\mathcal{M}) \to V(\mathcal{M})$$

with "$\cdot$" a place holder.

Note. It is easy to show that ∇Y is a $(1,1)$-tensor.

Consider two vector fields $X, Y \in V(\mathcal{M})$. Under coordinate basis $\{e_\mu = \partial_\mu\}$, we find

$$\nabla_X Y = X^\nu \nabla_{e_\nu}(Y^\mu e_\mu) = X^\nu(\partial_\nu Y^\mu + \Gamma^{\mu}_{\rho\nu} Y^\rho) e_\mu = X^\nu (\nabla Y)^\mu{}_\nu e_\mu$$

where

$$(\nabla Y)^\mu{}_\nu := \partial_\nu Y^\mu + \Gamma^{\mu}_{\rho\nu} Y^\rho \tag{25}$$

are identified as the components of ∇Y.

Note. We often denote this as $(\nabla Y)^\mu{}_\nu = (\nabla_\nu Y)^\mu = \nabla_\nu Y^\mu$. For further convenience, we sometimes use punctuation notations such as

$$Y^\mu{}_{,\nu} := \partial_\nu Y^\mu \quad \text{and} \quad Y^\mu{}_{;\nu} := \nabla_\nu Y^\mu$$

It is left as an exercise to check the components of ∇Y transform indeed as a $(1,1)$-tensor.

We succeeded in constructing the covariant derivative of vectors. How about covectors? We can define them by requiring the Leibniz rule.

Definition 3.14 (Covariant derivative of covectors). On manifold $\mathcal{M}$ with connection ∇, the covariant derivative of $\eta \in \Lambda^1(\mathcal{M})$ is a map

$$\nabla\eta : V(\mathcal{M}) \times V(\mathcal{M}) \to C^\infty(\mathcal{M})$$

$$(X, Y) \mapsto (\nabla_X \eta)(Y) := X(\eta(Y)) - \eta(\nabla_X Y) \tag{26}$$

for $X, Y \in V(\mathcal{M})$.

NOTE. It is also easy to verify that $\nabla\eta$ is a $(0,2)$-tensor.

Under coordinate basis $\{e_\mu = \partial_\mu\}$ for vectors and $\{f^\mu = \mathrm{d}x^\mu\}$ for covectors, we find the change in f^μ is

$$\nabla_\nu f^\mu = -\Gamma^\mu_{\nu\rho} f^\rho$$

by plugging basis vectors/covectors to relation (26). Exercise: check this.

With the above result, $\nabla\eta$ has components

$$(\nabla\eta)_{\mu\nu} := \partial_\nu \eta_\mu - \Gamma^\rho_{\mu\nu}\eta_\rho \tag{27}$$

Now, with covariant derivative of vectors and covectors in hand, we can generalise it to tensors by imposing the Leibniz rule over tensor products. We have the following inductive definition.

DEFINITION 3.15 (Covariant derivative of tensors). On manifold M with connection ∇, for general tensors S, T with covariant derivatives $\nabla S, \nabla T$, respectively, the covariant derivative of their tensor product is

$$\nabla(S \otimes T) := (\nabla S) \otimes T + S \otimes (\nabla T) \tag{28}$$

NOTE. This can be used as induction back to the base cases as functions, vectors and covectors. Also it is easy to check, for an (r,s)-tensor T, ∇T is a tensor of rank $(r, s+1)$.

Use the results we have, it is straightforward to show for an (r,s)-tensor T, the components of ∇T are

$$\begin{aligned}\nabla_\rho T^{\mu_1\cdots\mu_r}{}_{\nu_1\cdots\nu_s} &= \partial_\rho T^{\mu_1\cdots\mu_r}{}_{\nu_1\cdots\nu_s} \\ &\underbrace{+\Gamma^{\mu_1}_{\sigma\rho} T^{\sigma\mu_2\cdots\mu_r}{}_{\nu_1\cdots\nu_s} + \cdots + \Gamma^{\mu_r}_{\sigma\rho} T^{\mu_1\cdots\mu_{r-1}\sigma}{}_{\nu_1\cdots\nu_s}}_{\text{for each upstair index}} \\ &\underbrace{-\Gamma^{\sigma}_{\nu_1\rho} T^{\mu_1\cdots\mu_r}{}_{\sigma\nu_2\cdots\nu_s} - \cdots - \Gamma^{\sigma}_{\nu_s\rho} T^{\mu_1\cdots\mu_r}{}_{\nu_1\cdots\nu_{s-1}\sigma}}_{\text{for each downstair index}}\end{aligned} \tag{29}$$

EXERCISE. Prove the Leibniz rule of covariant derivatives holds in such way:

$$\begin{aligned}\nabla_X[T(\eta_1,\cdots,\eta_r;Y_1,\cdots,Y_s)] = \\ &(\nabla_X T)(\eta_1,\cdots,\eta_r;Y_1,\cdots,Y_s) \\ &+ T(\nabla_X\eta_1,\cdots,\eta_r;Y_1,\cdots,Y_s) + \cdots + T(\eta_1,\cdots,\nabla_X\eta_r;Y_1,\cdots,Y_s) \\ &+ T(\eta_1,\cdots,\eta_r;\nabla_X Y_1,\cdots,Y_s) + \cdots + T(\eta_1,\cdots,\eta_r;Y_1,\cdots,\nabla_X Y_s)\end{aligned}$$

To discuss further about connection, we need to introduce an object called torsion. Here we would not use the more formal definition as it is beyond our concern.

DEFINITION 3.16 (Torsion). On manifold $\mathcal{M}$ with connection ∇, the torsion is a $(1,2)$-tensor T with components

$$T_{\mu\nu}{}^\lambda = \Gamma^\lambda_{\mu\nu} - \Gamma^\lambda_{\nu\mu}$$

under coordinate basis $\{\partial_\mu\}$, where $\Gamma^\lambda_{\mu\nu}$ are the coefficients of ∇.

3.5 Levi-Civita Connection

Now we've discussed enough about metric and connection on their own, we want to see what happens when we put them together. The metric, equipped by a specific manifold, is the manifestation of its intrinsic properties, while, a connection can be arbitrarily chosen. The good news is, if we have a metric, then it gives us a unique connection for free.

THEOREM 3.4. *For a manifold $\mathcal{M}$ with metric g, there exists a unique, torsion free connection ∇ which is compatible with g, i.e.*

$$\nabla_X g = 0 \quad \forall X \in V(\mathcal{M})$$

This is the Levi-Civita connection.

Proof. We assume such connection ∇ exists. It is torsion free, i.e., the connection coefficients are symmetric in two lower indices: $\Gamma^\rho_{\mu\nu} = \Gamma^\rho_{\nu\mu}$. Imposing the condition $\nabla_X g = 0$, we find, under coordinate basis

$$\partial_\mu g_{\nu\rho} - \Gamma^\sigma_{\nu\mu} g_{\sigma\rho} - \Gamma^\sigma_{\rho\mu} g_{\nu\sigma} = 0 \tag{P1.1}$$

Permute (ρ, μ, ν) in a cyclic manner:

$$\partial_\nu g_{\rho\mu} - \Gamma^\sigma_{\rho\nu} g_{\sigma\mu} - \Gamma^\sigma_{\mu\nu} g_{\rho\sigma} = 0 \tag{P1.2}$$

$$\partial_\rho g_{\mu\nu} - \Gamma^\sigma_{\mu\rho} g_{\sigma\nu} - \Gamma^\sigma_{\nu\rho} g_{\mu\sigma} = 0 \tag{P1.3}$$

Use (P1.1) + (P1.2) − (P1.3), and by the torsion free property and symmetry of the metric, we get

$$(\partial_\mu g_{\nu\rho} + \partial_\nu g_{\rho\mu} - \partial_\rho g_{\mu\nu}) - 2\Gamma^\sigma_{\mu\nu} g_{\rho\sigma} = 0$$

Then by contracting with the inverse metric, we have

$$\boxed{\Gamma^\sigma_{\mu\nu} = \frac{1}{2} g^{\sigma\rho} (\partial_\mu g_{\nu\rho} + \partial_\nu g_{\rho\mu} - \partial_\rho g_{\mu\nu})} \tag{30}$$

These connection coefficients are unique by the properties of the metric.

Conversely, given such coefficients, it can be shown that they transform as (24). They are coefficients of a valid connection ∇. (It can be shown more formally, but out of our current concern.) It is easy to check this ∇ is torsion-free and metric compatible. □

NOTE. It is easy to show using Levi-Civita connection, $\nabla_X \hat{g} = 0$, $\forall X \in V(\mathcal{M})$ for the inverse metric $\hat{g}$.

Having Levi-Civita connection, we can

- Pull $g_{\mu\nu}$ or $g^{\mu\nu}$ out of covariant derivatives as constants and we can raise the indices of covariant derivatives by
$$\nabla^\mu := g^{\mu\nu} \nabla_\nu$$
- Under normal coordinates at $p \in \mathcal{M}$, we happily find that the connection coefficients vanish. And
$$\nabla_\mu \Big|_p = \partial_\mu \Big|_p$$
This is helpful when proving some tensorial identities. If certain property holds in normal coordinates, it should hold in any coordinates. We will see more on this later.

3.6 Some Useful Concepts

After the discussion of metric, connection and the Levi-Civita connection that a metric gives us for free, we can apply the covariant derivatives to get some useful concepts.

DEFINITION 3.17 (Divergence of vectors). The divergence of a vector $X \in V(\mathcal{M})$ is a scalar

$$\text{div } X = \nabla_\mu X^\mu$$

CLAIM 3.5.

$$\nabla_\mu X^\mu = \frac{1}{\sqrt{|g|}}\partial_\mu\left(\sqrt{|g|}X^\mu\right)$$

Proof.

$$\nabla_\mu X^\mu = \partial_\mu X^\mu + \Gamma^\mu_{\rho\mu}X^\rho$$

Consider

$$\partial_\mu g_{\nu\sigma} = \Gamma^\lambda_{\nu\mu}g_{\lambda\sigma} + \Gamma^\lambda_{\sigma\mu}g_{\nu\lambda}$$

Contract with $g^{\nu\sigma}$, we have

$$g^{\nu\sigma}\partial_\mu g_{\nu\sigma} = 2\Gamma^\lambda_{\lambda\mu}$$

Use the general result for a symmetric matrix $\mathbf{M}$ that

$$\ln(\det\mathbf{M}) = \text{Tr}(\ln\mathbf{M})$$

and differentiate, we have

$$\frac{1}{\det\mathbf{M}}\partial_\mu(\det\mathbf{M}) = \text{Tr}\big(\mathbf{M}^{-1}\partial_\mu\mathbf{M}\big)$$

which can be easily shown under a diagonalised basis.

Apply the above results to $g_{\mu\nu}$, we have

$$\Gamma^\mu_{\mu\rho} = \frac{1}{2g}\partial_\rho g = \frac{1}{\sqrt{|g|}}\partial_\rho\left(\sqrt{|g|}\right)$$

Then

$$\nabla_\mu X^\mu = \partial_\mu X^\mu + \frac{1}{\sqrt{|g|}}\partial_\mu\left(\sqrt{|g|}\right)X^\mu = \frac{1}{\sqrt{|g|}}\partial_\mu\left(\sqrt{|g|}X^\mu\right)$$

□

The divergence generalises to tensors, e.g. $\nabla_\mu T^{\mu\nu}$.

DEFINITION 3.18 (Curl of covectors). The curl of a covector $\eta \in \Lambda^1(\mathcal{M})$ is a $(0,2)$-tensor with components

$$(\text{curl } \eta)_{\mu\nu} = \nabla_\mu\eta_\nu - \nabla_\nu\eta_\mu$$

NOTE. It is easy to show

$$\nabla_\mu\eta_\nu - \nabla_\nu\eta_\mu = \partial_\mu\eta_\nu - \partial_\nu\eta_\mu$$

Definition 3.19 (Laplacian of scalars). The Laplacian of a scalar $f \in C^\infty(\mathcal{M})$ is a scalar

$$\nabla^2 f = \nabla^\mu \nabla_\mu f$$

Claim 3.6.

$$\nabla^\mu \nabla_\mu f = \frac{1}{\sqrt{|g|}} \partial_\mu \left(\sqrt{|g|} \partial^\mu f \right)$$

where we often denote $\partial^\mu = g^{\mu\nu} \partial_\nu$.

Proof. Left as an exercise. □

The Laplacian also generalises to tensors, e.g. $\nabla^2 T^{\mu\nu} = \nabla^\rho \nabla_\rho T^{\mu\nu}$.

Before we go on to discuss parallel transport and geodesics, we have one more thing left to introduce. It is actually an alternative notation for something we've just met.

Definition 3.20 (Intrinsic derivative). On a manifold $\mathcal{M}$ with connection ∇, the intrinsic derivative of a tensor T along a curve $\lambda(t)$ is

$$\frac{\mathrm{D}T}{\mathrm{D}t} := \nabla_X T$$

where $X = \frac{\mathrm{d}}{\mathrm{d}t}$ is the tangent vector of curve $\lambda(t)$.

With all these, we can start to discuss parallel transport and geodesics, which reflect more physical intuitions.

4 Parallel Transport and Geodesics

In this section, we introduce parallel transport of tensors and the geodesics on the manifold, based on all we have discussed about Differential Geometry. These two mathematical concepts are closely related. Later, they will play important roles when we investigate the physics of GR.

From now on, unless otherwise stated, all manifolds we use are equipped with a metric g and we use the Levi-Civita connection.

4.1 Parallel Transport

In Euclidean space, one nice and simple thing we can always do is translation of vectors. Now, when we have a general manifold, as vectors at different points live in different mathematical spaces, we cannot just brutally translate things around. Fortunately, we had an object called connection that surely "connects" tensors at different points. We can carefully generalise the notion of translation into an operation called parallel transport.

> Definition 4.1 (Parallel transport). On manifold $\mathcal{M}$, a tensor T is said to be parallel transported along curve $\lambda(u)$ if
> $$\frac{\mathrm{D}T}{\mathrm{D}u} = 0$$
> along $\lambda(u)$.

Parallel transport has the following properties:

(1) By ODE theory, the parallel transport equation defines the parallel transported tensor uniquely along the curve, provided the initial value is given.

(2) For a vector Y parallel transported along curve $\lambda(u)$, the operation is independent of parameterisation of the curve. This can be shown by
$$\delta Y^\mu = \frac{\mathrm{d}Y^\mu}{\mathrm{d}u}\delta u = -\Gamma^\mu_{\nu\rho} Y^\nu \frac{\mathrm{d}x^\rho}{\mathrm{d}u}\delta u = -\Gamma^\mu_{\nu\rho} Y^\nu \delta x^\rho$$
This argument can be generalised to tensors.

(3) The norm of a vector is preserved. For a vector Y parallel transported along curve $\lambda(u)$, we find
$$\frac{\mathrm{d}|Y|^2}{\mathrm{d}u} = \frac{\mathrm{D}}{\mathrm{D}u}(g_{\mu\nu}Y^\mu Y^\nu) = 2g_{\mu\nu}Y^\mu \frac{\mathrm{D}Y^\nu}{\mathrm{D}u} = 0$$

(4) The scalar product of two vectors parallel transported along the same curve is preserved. Exercise: show this.

In general, parallel transport is path dependent as a vector parallel transported along a closed loop is generally different from the original vector.

Example. On a 2-sphere described by polar coordinates (θ, ϕ), we take a path $(\theta = 0) \to (\theta = \pi/2, \phi = 0) \to (\theta = \pi/2, \phi = \pi/2) \to (\theta = 0)$. It can be shown that vectors parallel transported along this closed loop is rotated by an angle of $\pi/2$.

This path dependence is a manifestation of the intrinsic curvature of the manifold. We will discuss this in detail later in the course.

4.2 Interlude: Euler-Lagrange Equations

For us to study geodesics, we need a generic method using which we can find the functions that extremise a certain type of integrals.

Consider an integral

$$\mathcal{S}[x^\mu(u)] = \int_{u_1}^{u_2} \mathcal{L}(x^\mu, \dot{x}^\mu; u)\,\mathrm{d}u$$

where the integrand $\mathcal{L}$, always called *Lagrangian*, is a function of functions $x^\mu(u)$, $\dot{x}^\mu(u) = \frac{\mathrm{d}x^\mu}{\mathrm{d}u}$ and possibly the parameter u itself.

To extremise the integral, we can find the conditions $\mathcal{L}$ must satisfy. For a small variation $\delta x^\mu(u)$ of the function $x^\mu(u)$ with $\delta x^\mu(u_1) = \delta x^\mu(u_2) = 0$, the first order variation of $\mathcal{S}$ is

$$\begin{aligned}
\delta\mathcal{S} &= \int_{u_1}^{u_2} \left(\delta x^\mu \frac{\partial\mathcal{L}}{\partial x^\mu} + \delta\dot{x}^\mu \frac{\partial\mathcal{L}}{\partial\dot{x}^\mu}\right)\mathrm{d}u \\
&= \left[\delta x^\mu \frac{\partial\mathcal{L}}{\partial\dot{x}^\mu}\right]_{u_1}^{u_2} + \int_{u_1}^{u_2} \delta x^\mu \left[\frac{\partial\mathcal{L}}{\partial x^\mu} - \frac{\mathrm{d}}{\mathrm{d}u}\left(\frac{\partial\mathcal{L}}{\partial\dot{x}^\mu}\right)\right]\mathrm{d}u \\
&= \int_{u_1}^{u_2} \delta x^\mu \left[\frac{\partial\mathcal{L}}{\partial x^\mu} - \frac{\mathrm{d}}{\mathrm{d}u}\left(\frac{\partial\mathcal{L}}{\partial\dot{x}^\mu}\right)\right]\mathrm{d}u
\end{aligned}$$

As we want $\mathcal{S}$ to be at extremum, $\delta\mathcal{S} = 0$, the equations that should be satisfied by $\mathcal{L}$ are

$$\boxed{\frac{\mathrm{d}}{\mathrm{d}u}\left(\frac{\partial\mathcal{L}}{\partial\dot{x}^\mu}\right) = \frac{\partial\mathcal{L}}{\partial x^\mu}} \tag{31}$$

These equations are called the *Euler-Lagrange equations*. In fact, these equations are differential equations of functions $x^\mu(u)$.

4.3 Geodesics

In our discussion of geodesics, we only focus on Lorentzian manifold $(\mathcal{M}, g)$.

As there is a metric, we can calculate the length of a curve on the manifold. But the more interesting question is: given two points $p, q \in \mathcal{M}$, what is the shortest, or to be cautious, the **extreme** curve between them? In this subsection, we will investigate such curves, known as *geodesics*. They have key physical significance, as we will see later.

In the previous discussion, we've address the idea of a tangent vector of a curve at some point for many times. Here we make it a formal definition.

> DEFINITION 4.2 (Tangent vector of a curve). On manifold $\mathcal{M}$, the tangent vector (field) of a curve $\lambda(u)$ is
>
> $$X = \frac{\mathrm{d}}{\mathrm{d}u} = \frac{\mathrm{d}x^\mu}{\mathrm{d}u}\partial_\mu$$
>
> under coordinates $\{x^\mu\}$.

Recall the different types of vectors in a Lorentzian manifold, defined by their self scalar products. We can use the property of the tangent vectors to define the properties of curves.

DEFINITION 4.3 (Properties of curves). A curve is timelike/null/spacelike at a point $p \in \mathcal{M}$ iff its tangent vector at p is timelike/null/spacelike.

It is clear that for a non-null curve $\lambda(u)$, the norm of the tangent vector X is

$$|X| = |g_{\mu\nu}X^\mu X^\nu|^{1/2} = \left|\frac{\mathrm{d}s}{\mathrm{d}u}\right|$$

where s is the path length from the metric $\mathrm{d}s^2 = g_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu$.

Now, using the Lagrangian method, we would like to extremise the length of a **non-null** curve $\lambda(u)$ on a Lorentzian manifold $(\mathcal{M}, g)$ between two points $p = \lambda(0)$ and $q = \lambda(1)$.

The length of the selected segment can be written as

$$\mathcal{S} = \int_0^1 |g_{\mu\nu}\dot{x}^\mu \dot{x}^\nu|^{1/2}\,\mathrm{d}u$$

where $\dot{x}^\mu = \frac{\mathrm{d}x^\mu}{\mathrm{d}u}$, and the Lagrangian is identified as $\mathcal{L} = |g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu|^{1/2}$.

In fact, it is easy to see that the Lagrangian is simply the norm of the tangent vector under parameterisation u, i.e. $\mathcal{L} = \left|\frac{\mathrm{d}s}{\mathrm{d}u}\right|$, showing that the integral is invariant under re-parameterisation by chain rule.

We now solve the Euler-Lagrange equation to get the extreme curve between p and q. The functions we are looking for are the coordinates $x^\mu(u)$ for the extreme curve.

Euler-Lagrange equation is

$$\frac{\partial \mathcal{L}}{\partial x^\mu} = \frac{\mathrm{d}}{\mathrm{d}u}\left(\frac{\partial \mathcal{L}}{\partial \dot{x}^\mu}\right)$$

We calculate

$$\frac{\partial \mathcal{L}}{\partial x^\mu} = \pm\frac{1}{2\mathcal{L}}(\partial_\mu g_{\nu\rho})\dot{x}^\nu\dot{x}^\rho \quad \text{and} \quad \frac{\partial \mathcal{L}}{\partial \dot{x}^\mu} = \pm\frac{1}{\mathcal{L}}g_{\mu\nu}\dot{x}^\nu$$

where we take + for timelike curve and − for spacelike curve. Plug into the E-L equation, we get

$$\frac{\mathrm{d}}{\mathrm{d}u}\left(\frac{1}{\mathcal{L}}g_{\mu\nu}\dot{x}^\nu\right) = \frac{1}{2\mathcal{L}}(\partial_\mu g_{\nu\rho})\dot{x}^\nu\dot{x}^\rho$$

Then expand and calculate:

$$-\frac{1}{\mathcal{L}^2}\frac{\mathrm{d}\mathcal{L}}{\mathrm{d}u}g_{\mu\nu}\dot{x}^\nu + \frac{1}{\mathcal{L}}\frac{\mathrm{d}g_{\mu\nu}}{\mathrm{d}u}\dot{x}^\nu + \frac{1}{\mathcal{L}}g_{\mu\nu}\ddot{x}^\nu = \frac{1}{2\mathcal{L}}(\partial_\mu g_{\nu\rho})\dot{x}^\nu\dot{x}^\rho$$

$$-\frac{1}{\mathcal{L}}\ddot{s}g_{\mu\nu}\dot{x}^\nu + (\partial_\rho g_{\mu\nu})\dot{x}^\nu\dot{x}^\rho + g_{\mu\nu}\ddot{x}^\nu = \frac{1}{2}(\partial_\mu g_{\nu\rho})\dot{x}^\nu\dot{x}^\rho$$

$$g_{\mu\nu}\ddot{x}^\nu + \frac{1}{2}(\partial_\rho g_{\mu\nu} + \partial_\nu g_{\mu\rho} - \partial_\mu g_{\nu\rho})\dot{x}^\nu\dot{x}^\rho = \frac{\ddot{s}}{\dot{s}}g_{\mu\nu}\dot{x}^\nu$$

$$\ddot{x}^\sigma + \frac{1}{2}g^{\sigma\mu}(\partial_\rho g_{\mu\nu} + \partial_\nu g_{\mu\rho} - \partial_\mu g_{\nu\rho})\dot{x}^\nu\dot{x}^\rho = \frac{\ddot{s}}{\dot{s}}\dot{x}^\sigma$$

Finally, we identify the Levi-Civita connection coefficients $\Gamma^\sigma_{\nu\rho}$ and get

$$\ddot{x}^\sigma + \Gamma^\sigma_{\nu\rho}\dot{x}^\nu\dot{x}^\rho = \left(\frac{\ddot{s}}{\dot{s}}\right)\dot{x}^\sigma \tag{32}$$

where we use $\mathcal{L} = \pm\dot{s}$. These are the *geodesic equations.*

There is a preferred class of parameterisations that make our life easier. The condition is simply $\ddot{s} = 0$.

> Definition 4.4 (Affine parameter). On manifold $\mathcal{M}$, the parameter u of a curve $\lambda(u)$ is affine iff it is related to the path length s linearly, i.e.
>
> $$u = as + b$$
>
> where a, b are constants.

Using affine parameter, we get the simpler version of the geodesic equations, which are preferred.

$$\boxed{\ddot{x}^\mu + \Gamma^\mu_{\nu\rho}\dot{x}^\nu\dot{x}^\rho = 0} \tag{33}$$

This relates to the parallel transport by

$$\frac{\mathrm{D}X^\mu}{\mathrm{D}u} = 0$$

meaning that the tangent vector of the affinely parameterised geodesic is preserved along the curve.

Under affine parameterisation, we get another useful condition that

$$\mathcal{L} = \text{constant}$$

which turns out to be extremely useful.

Later when we discuss GR, we always use an affine parameter for convenience and for physical preference, as the affine parameters are linearly related to proper time/proper distance, which we will then define. For a non-null curve, the simplest choice of affine parameter is the path length s itself.

Since the original Lagrangian involves a square root, it is not convenient for calculation. Alternatively, with an affine parameter, we can use Lagrangian

$$\hat{\mathcal{L}} = g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu$$

to get the same result. Exercise: check this.

Using the alternative Lagrangian, we can compute the Levi-Civita connection coefficients a lot easier by directly solving the Euler-Lagrange equation and matching coefficients.

It is clear from ODE theory that, if we specify the initial conditions $x^\mu(0), \dot{x}^\mu(0)$, we can uniquely determine a geodesic $x^\mu(u)$.

We've now discussed enough about non-null geodesics heuristically. How about null curves? As they have $\mathrm{d}s^2 = 0$, we cannot use the Lagrangian method to extremise the length. We hereby unify the definition general geodesics using parallel transport.

DEFINITION 4.5 (Geodesic). An affinely parameterised geodesic is the integral curve of a vector field X which obeys

$$\nabla_X X = 0$$

The above definition has a profound physical significance. Say we take a curve $\lambda(\tau)$ parameterised by proper time τ. Take its tangent vector $U = \frac{\mathrm{d}}{\mathrm{d}\tau}$. Later we will see the components $U^\mu = \frac{\mathrm{d}x^\mu}{\mathrm{d}\tau}$ is actually the four-velocity of a particle. If we specify the initial position and four-velocity for some particle, then the curve is a geodesic if $\nabla_U U = 0$. Also, we would see that the four-acceleration is $\frac{\mathrm{D}U^\mu}{\mathrm{D}\tau} = (\nabla_U U)^\mu$. For a free particle, the four-acceleration should be zero. This suggests: the geodesics are trajectories followed by free particles.

5 Special Relativity Yet Again

Recall section 1 where we used elementary thought experiments to predict time dilation, length contraction and derive Lorentz transformations. Now, with adequate mathematical language, we formulate it again in a top-down manner, which gives us more clarity in this topic.

5.1 Lorentz Transformation

Here we will see Lorentz transformations again, however, we will properly define them this time.

Definition 5.1 (Minkowski spacetime). Minkowski spacetime is a four-dimensional Lorentzian manifold $(\mathcal{M}, \eta)$ with metric

$$\mathrm{d}s^2 = \eta_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu, \quad \eta_{\mu\nu} = \mathrm{diag}(+1, -1, -1, -1)$$

under global **Cartesian** coordinates $\{x^\mu\}$ in **inertial frames**.

Note. The global Cartesian coordinates $\{x^\mu\}$ we use have $\mu = 0, 1, 2, 3$ such that $x^0 = ct, x^1 = x, x^2 = y, x^3 = z$. Also it is easy to check that the inverse metric $\eta^{\mu\nu} = \mathrm{diag}(+1, -1, -1, -1)$ and connection coefficients $\Gamma^\mu_{\nu\rho} = 0$.

Definition 5.2 (Lorentz transformation). Lorentz transformation is a coordinate transformation between Cartesian coordinates $x^\mu \to x'^\mu$ that makes the metric unchanged, i.e.

$$\eta'_{\mu\nu} = \frac{\partial x^\rho}{\partial x'^\mu}\frac{\partial x^\sigma}{\partial x'^\nu}\eta_{\rho\sigma} = \eta_{\mu\nu}$$

Claim 5.1. *Lorentz transformation is linear so that we can write* $x'^\mu = \Lambda^\mu{}_\nu x^\nu + a^\mu$, *where* $\Lambda^\mu{}_\nu, a^\mu$ *are constants.*

Proof. Use

$$\eta_{\mu\nu} = \frac{\partial x'^\alpha}{\partial x^\mu}\frac{\partial x'^\beta}{\partial x^\nu}\eta_{\alpha\beta}$$

Take partial derivative with respect to x^ρ to get

$$\partial_\rho(\eta_{\mu\nu}) = \partial_\rho({x'^\alpha}_{,\mu}{x'^\beta}_{,\nu}\eta_{\alpha\beta})$$

$$({x'^\alpha}_{,\mu\rho}{x'^\beta}_{,\nu} + {x'^\alpha}_{,\mu}{x'^\beta}_{,\nu\rho})\eta_{\alpha\beta} = 0 \tag{P2.1}$$

Take even permutation of (μ, ν, ρ), we have

$$({x'^\alpha}_{,\nu\mu}{x'^\beta}_{,\rho} + {x'^\alpha}_{,\nu}{x'^\beta}_{,\rho\mu})\eta_{\alpha\beta} = 0 \tag{P2.2}$$

$$({x'^\alpha}_{,\rho\nu}{x'^\beta}_{,\mu} + {x'^\alpha}_{,\rho}{x'^\beta}_{,\mu\nu})\eta_{\alpha\beta} = 0 \tag{P2.3}$$

(P2.1) + (P2.2) − (P2.3) gives

$$2\frac{\partial x'^\alpha}{\partial x^\nu}\frac{\partial^2 x'^\beta}{\partial x^\mu \partial x^\rho}\eta_{\alpha\beta} = 0$$

Since $\frac{\partial x'^{\alpha}}{\partial x^{\nu}} \neq 0$ in general, we have

$$\frac{\partial^2 x'^{\beta}}{\partial x^{\mu}\partial x^{\rho}} = 0$$

which suggests $x^{\mu} \to x'^{\mu}$ is linear and there exists constants $\Lambda^{\mu}{}_{\nu}, a^{\mu}$ such that

$$\boxed{x'^{\mu} = \Lambda^{\mu}{}_{\nu} x^{\nu} + a^{\mu}} \tag{34}$$

□

> DEFINITION 5.3 (Homogeneous/inhomogeneous Lorentz transformation). For a Lorentz transformation $x'^{\mu} = \Lambda^{\mu}{}_{\nu} x^{\nu} + a^{\mu}$, if $a^{\mu} = 0$, it is called homogeneous Lorentz transformation (or just Lorentz transformation); if $a^{\mu} \neq 0$, translation presents, it is called inhomogeneous Lorentz transformation or Poincaré transformation.

Now we have the requirement for a general Lorentz transformation by identifying $\Lambda^{\mu}{}_{\nu} = \frac{\partial x'^{\mu}}{\partial x^{\nu}}$,

$$\eta_{\mu\nu} = \Lambda^{\rho}{}_{\mu}\Lambda^{\sigma}{}_{\nu}\eta_{\rho\sigma} \tag{†}$$

From this, it is easy to find that the inverse Lorentz transformation can be obtained by raising/lowering the downstairs/upstairs indices

$$(\Lambda^{-1})^{\nu}{}_{\mu} = \Lambda_{\mu}{}^{\nu} = \eta_{\rho\mu}\eta^{\sigma\nu}\Lambda^{\rho}{}_{\sigma}$$

Also from (†), it is clear that

(1) $[\det(\Lambda^{\mu}{}_{\nu})]^2 = 1$,

(2) set $\mu = \nu = 0$, we get $(\Lambda^0{}_0)^2 - \sum_{i=1}^{3}(\Lambda^i{}_0)^2 = 1$, i.e. $(\Lambda^0{}_0)^2 \geq 1$.

> DEFINITION 5.4 (Proper Lorentz transformation). Proper Lorentz transformation is a homogeneous Lorentz transformation with conditions
>
> (i) $\det(\Lambda^{\mu}{}_{\nu}) = 1$, i.e. keeps the same spatial handedness;
>
> (ii) $\Lambda^0{}_0 \geq 1$, i.e. there is no time reversal.

In our course, we only consider proper Lorentz transformation.

Now we can explore the classes of solutions to $\Lambda^{\mu}{}_{\nu}$ obeying (†).

(1) Rotation:

$$\Lambda^{\mu}{}_{\nu} = \left(\begin{array}{c|c} 1 & 0 \\ \hline 0 & \mathbf{R} \end{array}\right)$$

with $\mathbf{R}$ a spatial rotation is an obvious solution.

(2) Boost: from section 1, the boost with relative speed v along x-direction under standard configuration is

$$\Lambda^{\mu}{}_{\nu} = \left(\begin{array}{cc|cc} \gamma & -\gamma\beta & \multicolumn{2}{c}{0} \\ -\gamma\beta & \gamma & & \\ \hline \multicolumn{2}{c|}{0} & 1 & \\ & & & 1 \end{array}\right)$$

The general derivation of such solutions is fiddly and involves arbitrariness, we are not showing it here. But it is easy to check that the above $\Lambda^{\mu}{}_{\nu}$ all obey (†).

Now as coordinates transformation as $x^{\mu} \to x'^{\mu} = \Lambda^{\mu}{}_{\nu} x^{\nu}$, the coordinate basis transforms as

$$\partial_{\mu} \to \partial'_{\mu} = \frac{\partial x^{\nu}}{\partial x'^{\mu}} \partial_{\nu} = \Lambda_{\mu}{}^{\nu} \partial_{\nu}$$

Example. Simply consider 2D Minkowski spacetime with coordinates (ct, x). If

$$\Lambda^{\mu}{}_{\nu} = \begin{pmatrix} \gamma & -\gamma\beta \\ -\gamma\beta & \gamma \end{pmatrix}$$

it is easy to check

$$\Lambda_{\mu}{}^{\nu} = \begin{pmatrix} \gamma & \gamma\beta \\ \gamma\beta & \gamma \end{pmatrix}$$

and we have

$$e'_0 = \gamma(e_0 + \beta e_1)$$
$$e'_1 = \gamma(\beta e_0 + e_1)$$

for coordinate basis $\{e_{\mu} = \partial_{\mu}\}$.

As we have studied the most important coordinate transformations in SR, we can now consider vectors, covectors, and generally, tensors.

Definition 5.5 (4-vector). A 4-vector is a vector in four-dimensional spacetime.

It is easy to see the components of 4-vector X transform under $x^{\mu} \to x'^{\mu} = \Lambda^{\mu}{}_{\nu} x^{\nu}$ as

$$X'^{\mu} = \Lambda^{\mu}{}_{\nu} X^{\nu}$$

Similarly, we determine the property of a 4-vector X as timelike/null/spacelike as $\langle X, X\rangle > 0$ or $\langle X, X\rangle = 0$ or $\langle X, X\rangle < 0$. Lightcone of each point can be defined by all null 4-vectors at the point.

Example. Coordinate basis ∂_0 is timelike; $\partial_i, i = 1, 2, 3$ are spacelike.

Definition 5.6 (Future/past pointing 4-vector). A timelike/null 4-vector X is future pointing if $X^0 > 0$; it is past pointing if $x^0 < 0$.

We often write components of a 4-vector as a scalar component with a 3-vector component, e.g.

$$X^{\mu} = (X^0, \mathbf{X})$$

By the natural isomorphism established by metric $\eta_{\mu\nu}$, we can always have a covector with components

$$X_{\mu} = \eta_{\mu\nu} X^{\nu} = (X^0, -\mathbf{X})$$

generated from X^μ.

Clearly, components of any 4-covector ω transform as

$$\omega_\mu = \Lambda_\mu{}^\nu \omega_\nu$$

The transformational laws of a general tensor in Minkowski spacetime can be constructed on that of 4-vectors and 4-covectors.

5.2 Relativistic Dynamics

Before we dive into the discussion of dynamics of relativistic particles, we first reiterate the definition of proper time.

> Definition 5.7 (Proper time). The proper time τ is defined as $c^2 \mathrm{d}\tau^2 = \mathrm{d}s^2$ where $\mathrm{d}s^2$ is the squared infinitesimal path difference in the spacetime.

To describe the motion of any particle in the spacetime, we introduce the concept of a worldline.

> Definition 5.8 (Worldline). The worldline of a particle is its trajectory $x^\mu(u)$ in the spacetime, parameterised by some $u \in \mathbb{R}$.

5.2.1 Massive Particles

Here we first discuss massive particles (i.e. particles with non-zero mass) as they are more intuitive to deal with. Later, we will see that the physics of massless particles can be generalised from the notions developed when describing massive particles.

Kinematics

> Claim 5.2. *Worldline of a massive particle is always timelike.*

Proof. It is easy to see from the fact that the property of a curve in the spacetime is preserved under Lorentz transformation. We can always find the rest frame of any massive particle, in which the worldline is trivially timelike. Then in general this should hold. □

Note. The proper time τ is always a convenient parameter for timelike worldlines. It is affine.

> Definition 5.9 (4-velocity). The 4-velocity of a massive particle is the tangent vector U to the worldline with parameter τ.

4-velocity has components

$$U^\mu = \frac{\mathrm{d}x^\mu}{\mathrm{d}\tau}$$

This in general gives

$$\langle U, U \rangle = c^2 \tag{35}$$

as $U_\mu U^\mu = \left(\frac{\mathrm{d}s}{\mathrm{d}\tau}\right)^2 = c^2$.

We can write the 4-velocity of a massive particle in terms of speed of light c and the 3-velocity of the particle $\mathbf{u}$ as

$$U^\mu = \gamma_u(c, \mathbf{u})$$

where γ_u is the Lorentz factor obtained by the 3-speed $u = |\mathbf{u}|$.

With our mathematical language, it is more convenient to find the velocity transformational laws.

For two inertial frames S and S' with coordinates $\{x^\mu\}$ and $\{x'^\mu\}$, respectively, related by Lorentz transformation $x'^\mu = \Lambda^\mu{}_\nu x^\nu$ (relative speed v in x-direction, under standard configuration), then

$$U'^\mu = \Lambda^\mu{}_\nu U^\nu$$

In matrix notation,

$$\begin{pmatrix} \gamma_{u'} c \\ \gamma_{u'} u'^1 \\ \gamma_{u'} u'^2 \\ \gamma_{u'} u'^3 \end{pmatrix} = \begin{pmatrix} \gamma_v & -\beta\gamma_v & 0 & 0 \\ -\beta\gamma_v & \gamma_v & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \gamma_u c \\ \gamma_u u^1 \\ \gamma_u u^2 \\ \gamma_u u^3 \end{pmatrix}$$

where $u^i, i = 1, 2, 3$ are the components of 3-velocity $\mathbf{u}$, the same holds for $\mathbf{u}'$.

We then have

$$\frac{\gamma_u}{\gamma_{u'}} = \frac{1}{\gamma_v(1 - u^1 v/c^2)} \tag{36}$$

with which we further get

$$\begin{aligned} u'^1 &= \frac{u^1 - v}{1 - u^1 v/c^2} \\ u'^2 &= \frac{u^2}{\gamma_v(1 - u^1 v/c^2)} \\ u'^3 &= \frac{u^3}{\gamma_v(1 - u^1 v/c^2)} \end{aligned} \tag{37}$$

which agrees with our previous results from section 1.

Before we go on to define the acceleration, we first state a theorem.

THEOREM 5.3. *In an inertial frame, free massive particles move on timelike geodesics in Minkowski spacetime.*

Proof. By the definition of inertial frame, a free massive particle should have $\frac{\mathrm{d}^2 x^i}{\mathrm{d}t^2} = 0$ for $i = 1, 2, 3$ by Newton's first law. This suggests $\gamma, \mathbf{u}$ are constants, giving $U^\mu = \gamma(c, \mathbf{u})$ a constant. Then it should obey $\frac{\mathrm{d}U}{\mathrm{d}\tau} = 0$. Since it is in Minkowski spacetime, $\Gamma^\rho_{\mu\nu} = 0$ everywhere, we can say tensorially,

$$\frac{\mathrm{D}U}{\mathrm{D}\tau} = 0$$

This means the worldline $x^\mu(\tau)$ obeys

$$\ddot{x}^\mu + \Gamma^\mu_{\nu\rho}\dot{x}^\nu \dot{x}^\rho = 0$$

with affine parameter τ, by recalling that $U^\mu = \frac{\mathrm{d}x^\mu}{\mathrm{d}\tau}$. This shows that such worldline is a timelike geodesic.

The converse is also true, seen easily. □

As acceleration should distinct accelerated particles from free ones, we can see that the vector $\frac{\mathrm{D}U}{\mathrm{D}\tau}$ is of physical importance, and will be defined as the acceleration.

DEFINITION 5.10 (4-acceleration). The 4-acceleration of a massive particle with 4-velocity U and proper time τ is a vector G such that

$$G = \frac{\mathrm{D}U}{\mathrm{D}\tau} = \nabla_U U$$

NOTE. We immediately see that the above theorem states that *free particle* $\Leftrightarrow G = 0$, which generalises Newton's first law. In Minkowski spacetime, $G = \frac{\mathrm{d}U}{\mathrm{d}\tau}$ since $\Gamma^{\rho}_{\mu\nu} = 0$.

It is easy to show

$$\langle G, U \rangle = 0$$

as

$$G_\mu U^\mu = \frac{\mathrm{D}U_\mu}{\mathrm{D}\tau} U^\mu = \frac{1}{2}\frac{\mathrm{D}}{\mathrm{D}\tau}\left(U_\mu U^\mu\right) = 0$$

Write $U^\mu = \gamma_u(c, \mathbf{u})$, we have

$$G^\mu = \frac{\mathrm{d}U^\mu}{\mathrm{d}\tau} = \gamma_u \frac{\mathrm{d}}{\mathrm{d}t}(\gamma_u c, \gamma_u \mathbf{u})$$

Since

$$\frac{\mathrm{d}\gamma_u}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}\left(1 - \frac{\mathbf{u}\cdot\mathbf{u}}{c^2}\right)^{-1/2} = \frac{\gamma_u^3}{c^2}\mathbf{u}\cdot\mathbf{a}$$

where $\mathbf{a} = \frac{\mathrm{d}\mathbf{u}}{\mathrm{d}t}$ is the 3-acceleration. Then

$$G^\mu = \gamma_u^2 \left(\frac{\gamma_u^2}{c}\mathbf{u}\cdot\mathbf{a}, \mathbf{a} + \frac{\gamma_u^2}{c^2}(\mathbf{u}\cdot\mathbf{a})\mathbf{u}\right) \tag{38}$$

Recall the definition of instantaneous rest frame (IRF). In IRF, $\mathbf{u} = 0, \mathbf{a} = \mathbf{a}_{\mathrm{IRF}}$, this gives

$$G^\mu = (0, \mathbf{a}_{\mathrm{IRF}})$$

and in general,

$$\langle G, G \rangle = -|\mathbf{a}_{\mathrm{IRF}}|^2 < 0 \tag{39}$$

meaning that G is always spacelike.

Dynamics

To discuss the dynamics of relativistic massive particles, first we derive the 3-momentum $\mathbf{p}$ and energy E. Since the derivation of these are not elegant (and these should definitely be taught in any elementary SR course), we just quote the result:

$$\mathbf{p} = \gamma_u m \mathbf{u} \qquad E = \gamma_u m c^2 \tag{40}$$

for a massive particle of mass m travelling at 3-velocity $\mathbf{u}$.

To use the language of differential geometry, we have the following definition.

DEFINITION 5.11 (4-momentum). The 4-momentum of a massive particle with mass m is

$$P = mU$$

where U is its 4-velocity.

We identify

$$P^\mu = (\gamma_u mc, \gamma_u m\mathbf{u})$$

or

$$P^\mu = \left(\frac{E}{c}, \mathbf{p}\right)$$

Note that $\langle P, P\rangle = m^2c^2$ and we have the famous *energy-momentum invariant*

$$E^2 - |\mathbf{p}|^2c^2 = m^2c^4 \tag{41}$$

NOTE. For an isolated system of massive particles undergoing collision, the total 4-momentum can be written as the sum of all the individual (since Minkowski spacetime is pseudo-Euclidean) and is conserved. This combines energy and 3-momentum conservation together in a Lorentz invariant way.

Similarly, we can define the four-dimensional analogue of force.

DEFINITION 5.12 (4-force). The 4-force of a massive particle with 4-momentum P is

$$F = \frac{\mathrm{D}P}{\mathrm{D}\tau}$$

We identify

$$\boxed{F = mG} \tag{42}$$

for a particle with mass m, which is the generalised Newton's second law.

Writing the components of F explicitly:

$$F^\mu = \gamma_u \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{E}{c}, \mathbf{p}\right) = \gamma_u \left(\frac{\mathbf{f}\cdot\mathbf{u}}{c}, \mathbf{f}\right) \tag{43}$$

where $\mathbf{f}$ is the 3-force.

Similarly, we find

$$\langle F, U\rangle = 0$$

5.2.2 Massless Particles

We now turn to massless particles. These particles always have the speed of light and their worldlines are null, as we will shortly prove.

CLAIM 5.4. *Massless particles travel at speed of light c. Their worldlines are null.*

Proof. Assuming usual notations, consider the identity

$$\mathbf{u} = \frac{\gamma m\mathbf{u}}{\gamma m} = \frac{c^2\mathbf{p}}{E} = \frac{c^2\mathbf{p}}{\sqrt{m^2c^4 + |\mathbf{p}|^2c^2}}$$

For massless particles, $m = 0$, we have

$$\mathbf{u} = \frac{\mathbf{p}}{|\mathbf{p}|}c$$

which has $|\mathbf{u}| = c$.

Now, for Minkowski spacetime, the infinitesimal spacetime interval on their worldlines is

$$\mathrm{d}s^2 = c^2\mathrm{d}t^2 - \mathrm{d}x^2 - \mathrm{d}y^2 - \mathrm{d}z^2 = \mathrm{d}t^2\left(c^2 - |u|^2\right) = 0$$

For a general spacetime, we can always find normal coordinates at each point to see that the above holds. This is essentially coordinate independent. □

We now turn to momentum and energy for massless particles.

From E-p invariant, with $m = 0$, we have

$$E = |\mathbf{p}|c \tag{44}$$

DEFINITION 5.13 (4-momentum of massless particles). The 4-momentum of a massless particle with 3-momentum $\mathbf{p}$ is defined as

$$P^\mu = (|\mathbf{p}|, \mathbf{p})$$

NOTE. $\langle P, P\rangle = 0$ for massless particles.

THEOREM 5.5. *The worldlines of free massless particles are null geodesics, and can be affinely parameterised.*

Proof. For a free massless particle, its worldline with any valid parameter σ is $x^\mu(\sigma)$. Note that $\mathrm{d}\tau = 0$, we cannot use proper time τ as parameter here. Its 4-momentum P obeys

$$\frac{\mathrm{D}P}{\mathrm{D}\sigma} = 0$$

for any valid σ as for a free particle $\mathbf{p}$ is constant.

For $x^\mu(\sigma)$ to be an affinely parameterised geodesic, we need σ to satisfy

$$P^\mu \propto \frac{\mathrm{d}x^\mu}{\mathrm{d}\sigma}$$

First consider in some inertial frame, its coordinate time t can be used as a parameter since

$$P^\mu = \frac{E}{c}\left(1, \frac{\mathbf{p}}{|\mathbf{p}|}\right) = \frac{E}{c^2}\left(c\frac{\mathrm{d}t}{\mathrm{d}t}, \frac{\mathrm{d}x}{\mathrm{d}t}, \frac{\mathrm{d}y}{\mathrm{d}t}, \frac{\mathrm{d}z}{\mathrm{d}t}\right) = \frac{E}{c^2}\frac{\mathrm{d}x^\mu}{\mathrm{d}t}$$

which follows from the fact that massless particles travel at c.

Therefore we just choose σ such that

$$\frac{\mathrm{d}t}{\mathrm{d}\sigma} \propto \frac{E}{c^2}$$

which gives $P^\mu \propto \frac{\mathrm{d}x^\mu}{\mathrm{d}\sigma}$.

Thus $\frac{\mathrm{D}P}{\mathrm{D}\sigma} = 0$ is an affine null geodesic equation, $x^\mu(\sigma)$ is an affine null geodesic. □

As photons are massless particles, we can generalise wavevectors $\mathbf{k}$.

DEFINITION 5.14 (4-wavevector). The 4-wavevector for a photon with 4-momentum P is

$$K = \frac{P}{\hbar}$$

where $\hbar$ is the (reduced) Planck's constant from quantum theory.

We identify

$$K^\mu = (|\mathbf{k}|, \mathbf{k}) = \left(\frac{2\pi}{\lambda}, \mathbf{k}\right)$$

where $\mathbf{k}$ is the 3-momentum and λ is the wavelength.

EXAMPLE (Doppler effect). Two inertial frames S and S' are under standard configuration and S' is travelling at constant speed v in the x-direction. They can be related by a Lorentz transformation $\Lambda^\mu{}_\nu$.

For some photon, in S it is observed as $K^\mu = \frac{2\pi}{\lambda}(1, \cos\theta, \sin\theta, 0)$; in S', it has $K'^\mu = \Lambda^\mu{}_\nu K^\nu$.

Only consider the zeroth component as an example,

$$K'^0 = \frac{2\pi}{\lambda'} = \frac{2\pi}{\lambda}\gamma(1 - \beta\cos\theta)$$

we get

$$\frac{\lambda}{\lambda'} = \gamma(1 - \beta\cos\theta) \tag{45}$$

For $\theta = 0$, we recover the previous example where

$$\frac{\lambda}{\lambda'} = \sqrt{\frac{1-\beta}{1+\beta}}$$

EXERCISE. Find the change in frequency of a photon collided with an electron with angle change θ.

5.3 Local Reference Frame

Recall the definition of instantaneous reference frame in section 1. We here use the mathematics learned to describe what an IRF basis looks like for a general observer.

DEFINITION 5.15 (Orthogonal tetrad). An orthogonal tetrad $\{e_\mu\}$ are a basis to four-dimensional Minkowski spacetime such that

$$g(e_\mu, e_\nu) = \eta_{\mu\nu} \quad \text{and} \quad ce_0(\tau) = U(\tau)$$

where $U(\tau)$ is the 4-velocity of some massive particle at proper time τ.

CLAIM 5.6. *Such orthogonal tetrad is the basis of IRF for any massive particle with instantaneous 4-velocity $U(\tau)$ at τ.*

Proof. Use Lorentz transformation. □

5.4 A Sniff of GR

Let inertial frame S have coordinates $X^\mu = (cT, X, Y, Z)$.

Introduce new rotating frame S', under standard configuration with S, rotating about Z-axis at angular speed ω. Then for $x^\mu = (ct, x, y, z)$ in S', we have

$$\begin{aligned} X &= x\cos\omega t - y\sin\omega t \\ Y &= x\sin\omega t + y\cos\omega t \\ Z &= z \\ T &= t \end{aligned}$$

We can now see how the metric changes. In S,

$$\mathrm{d}s^2 = c^2\mathrm{d}T^2 - \mathrm{d}X^2 - \mathrm{d}Y^2 - \mathrm{d}Z^2$$

In S', after some tedious calculation,

$$\mathrm{d}s^2 = [c^2 - \omega^2(x^2+y^2)]\mathrm{d}t^2 + 2\omega y\,\mathrm{d}t\,\mathrm{d}x - 2\omega x\,\mathrm{d}t\,\mathrm{d}y - \mathrm{d}x^2 - \mathrm{d}y^2 - \mathrm{d}z^2$$

We further find the geodesics in x^μ coordinates are

$$\begin{aligned} \ddot{x} &= \omega^2 x + 2\omega\dot{y} \\ \ddot{y} &= \omega^2 y - 2\omega\dot{x} \\ \ddot{z} &= 0 \end{aligned}$$

with $\dot{} \equiv \frac{\mathrm{d}}{\mathrm{d}t}$.

Free particles don't follow "straight lines" in S': spacetime seems to be curved! This means acceleration somehow brings curvature into spacetime. Later we will discuss Equivalence Principle which states that acceleration and gravity are equivalent. This gives us a sniff of GR: gravity means curvature in spacetime!

6 An Interlude: Electromagnetism

In this section, I would like to show how we can use the language of SR to describe Electromagnetism. As we are limited in time, this section is only for a short description rather than for educational purposes. If we dive into the subject of EM too deep, it would take me at least 16 hours to tell you the whole story! So let's just have a peek and move on. If you are interested in this subject, please refer to other excellent notes an books on Electromagnetism or Electrodynamics.

6.1 Intuition

Magnetism is a purely relativistic effect. Consider a wire with the following property. When we are at the rest frame of it, S, the densities of positive and negative charges in it are equal, and they travel at opposite directions at the same speed v. The wire appears neutral in S, but there is a current. Now we have an test charge q with speed u with respect to the wire. In S, as there is a current, the moving test charge experiences a magnetic force. However, if we Lorentz transform to the rest frame of the test charge, S', the charge densities of positive and negative are different as the speeds they travel at differ (Lorentz contraction). There is a pure electric force appear in S' rather than magnetic, however, the physics should remain the same. This reveals the relativistic nature of Electromagnetism. In fact, Maxwell's equations are naturally relativistic (although they are not expressed covariantly).

6.2 General Plot

To describe the source of EM field, we start from charge density $\rho(\mathbf{x}, t)$.

DEFINITION 6.1 (4-current density). The 4-current density of charge density ρ is

$$J = \rho U$$

where U is the 4-velocity of the charge.

To describe the field, we combine the electric scalar potential and magnetic vector potential together.

DEFINITION 6.2 (4-vector potential). The 4-vector potential of the electromagnetic field is

$$A^\mu = \left(\frac{\phi}{c}, \mathbf{A}\right)$$

where ϕ is the electric potential and $\mathbf{A}$ is the magnetic vector potential.

From these we describe the electric field $\mathbf{E}$ and magnetic flux density $\mathbf{B}$ by

$$\mathbf{E} = -\boldsymbol{\nabla}\phi - \dot{\mathbf{A}} \qquad \mathbf{B} = \boldsymbol{\nabla} \times \mathbf{A}$$

where $\boldsymbol{\nabla} \equiv \frac{\partial}{\partial x^i}, i = 1, 2, 3$ and $\dot{} \equiv \frac{\partial}{\partial t}$.

Then we can use the language of SR to describe the field.

DEFINITION 6.3 (Field strength tensor). The field strength tensor $F_{\mu\nu}$ is defined by

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu = \begin{pmatrix} 0 & E^1/c & E^2/c & E^3/c \\ -E^1/c & 0 & -B^3 & B^2 \\ -E^2/c & B^3 & 0 & -B^1 \\ -E^3/c & -B^2 & B^1 & 0 \end{pmatrix}$$

where $E^i, B^i, i = 1, 2, 3$ are the components of electric field $\mathbf{E}$ and magnetic flux density $\mathbf{B}$.

The Lorentz transformation of $\mathbf{E}$ and $\mathbf{B}$ can be found by

$$F'^{\mu\nu} = \Lambda^\mu{}_\rho \Lambda^\nu{}_\sigma F^{\rho\sigma}$$

With these notions, we can write compactly

Continuity Equation

$$\dot{\rho} + \boldsymbol{\nabla} \cdot \mathbf{J} = 0 \qquad \Leftrightarrow \qquad \partial_\mu J^\mu = 0 \tag{46}$$

where $\mathbf{J}$ is the 3-current density.

Lorentz Force

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \qquad \Leftrightarrow \qquad f_\mu = qF_{\mu\nu}U^\nu \tag{47}$$

where q is the charge of the particle, f_μ is the 4-force.

Maxwell's Equations

$$\left.\begin{aligned} \boldsymbol{\nabla} \cdot \mathbf{E} &= \tfrac{\rho}{\epsilon_0} \\ \boldsymbol{\nabla} \times \mathbf{B} &= \mu_0 \mathbf{J} + \mu_0 \epsilon_0 \dot{\mathbf{E}} \end{aligned}\right\} \qquad \Leftrightarrow \qquad \partial_\mu F^{\mu\nu} = \mu_0 J^\nu \tag{48}$$

$$\left.\begin{aligned} \boldsymbol{\nabla} \cdot \mathbf{B} &= 0 \\ \boldsymbol{\nabla} \times \mathbf{E} &= -\dot{\mathbf{B}} \end{aligned}\right\} \qquad \Leftrightarrow \qquad \partial_{[\mu} F_{\nu\rho]} = 0 \tag{49}$$

In curved spacetime, we just have to change $\partial_\mu \to \nabla_\mu$, i.e.

$$\nabla_\mu F^{\mu\nu} = \mu_0 J^\nu \qquad \nabla_{[\mu} F_{\nu\rho]} = 0$$

To be continued...

Diagrams are under construction.

Thank you for reading for now.

References

[1] A. Challinor, *Part II Relativity*

[2] U. Sperhake, *Part II General Relativity*

[3] D. Tong, *Part III General Relativity*

[4] H.S. Reall, *Part 3 General Relativity*

[5] M.P. Hobson, G.P. Efstathion, A.N. Lasenby, *General Relativity: An Introduction for Physicists* (CUP 2006)