

GENERAL RELATIVITY & COSMOLOGY

A Quick Guide

Huan Bui

Colby College
Physics & Statistics
Class of 2021

February 4, 2019

Preface

Greetings,

General Relativity & Cosmology - A Quick Guide to is compiled from my PH335: General Relativity and Cosmology with professor Robert Bluhm at Colby College. The course is based on *A Short Course in General Relativity*, 3th Edition, by James Foster and J. David Nightingale.

One of my favorite things to do is to transfer my handwritten lecture notes into nice L^AT_EX documents. I started working on these documents last year with *Special Relativity - A Quick Guide to*, which was based on (part of) my PH241: Modern Physics I with professor Charles Conover. I am still working on it, along with another one on Linear Algebra (based on MA253 with Otto Bretscher), and of course General Relativity - the one you are reading now. All of these .pdf files can be found on my [website](#), under the “A Quick Guide to...” tab. For the “raw” content, i.e., my handwritten lecture notes, feel free to visit the “Lecture Notes” tab.

Today is Dec 22, 2018. In the spring I will be doing an independent study on Classical Fields Theory with prof. Bluhm. I will also be taking Matrix Analysis with prof. Leo Livshits. I’m excited to be taking these advanced physics and mathematics courses, and I look forward to sharing my journey with you through my lecture notes.

Enjoy!

Contents

Preface	1
1 Overview and Review	5
1.1 Review of Special Relativity	5
1.2 The Equivalence Principle	7
1.3 Versions of the Equivalence Principle	9
1.3.1 The Strong Equivalence Principle	9
1.3.2 The Weak Equivalence Principle	9
2 Review of Vector Calculus	10
2.1 Operations & Theorems	10
2.2 Coordinate Systems	14
3 Flat 3-dimensional space	17
3.1 Curvilinear coordinates	18
3.2 Basis vectors	18
3.3 Natural basis	18
3.4 Dual basis	21
3.5 Contravariant and covariant vectors	23
3.5.1 The suffix notation (vector calculus)	23
3.5.2 The Einstein summation notation	23
3.5.3 Contravariant and covariant vectors	24
3.6 Metric tensor	29
3.6.1 Summations as matrix products	31
3.7 Coordinate transformation	34
3.7.1 Vectors	34
3.7.2 Tensors	42
3.8 Scalars	44
4 Flat spacetime	45
4.1 Special Relativity	45
4.1.1 The Lorentz Transformations, revisited	47
4.1.2 A curiosity about the Lorentz boosts	50
4.1.3 The Poincaré transformations	51
4.1.4 Velocity, momentum, and force	55
4.2 Relativistic Electrodynamics	61
5 Curved spaces	65
5.1 Geodesics	65
5.2 2-dimensional curved spaces	66
5.3 Manifolds	70
5.4 Tensors on manifolds	71
5.4.1 Combining tensors	71
5.4.2 Special tensors	74

6	Gravitation and Curvature	75
6.1	Curvature	75
6.2	Geodesics and Affine connections $\Gamma_{\mu\nu}^{\sigma}$	76
6.2.1	Flat 3D space	76
6.3	Geodesics in curved space	81
6.4	Parallel transport	83
6.5	Curved Spacetime	86
6.6	Principle of covariance	86
6.7	Absolute and Covariant differentiation	89
6.7.1	Absolute differentiation	89
6.7.2	Covariant differentiation	92
6.8	Newtonian limit	94
6.8.1	Overview	94
6.8.2	Weak limit of General Relativity	96
7	Einstein's field equations	100
7.1	The stress-energy tensor $T^{\mu\nu}$	102
7.2	Riemann curvature tensor $R^{\lambda}_{\mu\nu\sigma}$	103
7.3	The Einstein equations	105
7.4	Schwarzschild solution	105
8	Predictions and tests of general relativity	106
8.1	Gravitational redshift	106
8.2	Radar time-delay experiments	106
8.3	Black Holes	106
9	Cosmology	107
9.1	Large-scale geometry of the universe	107
9.1.1	Cosmological principle	107
9.1.2	Robertson-Walker (flat, open, closed) geometries	107
9.1.3	Expansion of the universe	107
9.1.4	Distances and speeds	107
9.1.5	Redshifts	107
9.2	Dynamical evolution of the universe	107
9.2.1	The Friedmann equations	107
9.2.2	The cosmological constant Λ	107
9.2.3	Equations of state	107
9.2.4	A matter-dominated universe ($\Lambda = 0$) [Friedmann models]	107
9.2.5	A flat, matter-dominated universe ($\Lambda = 0$) [old favorite model]	107
9.3	Observational cosmology	107
9.3.1	Hubble law	107
9.3.2	Acceleration of the universe	107
9.3.3	Matter densities and dark matter	107
9.3.4	The flatness and horizon problems	107
9.3.5	Cosmic Microwave Background (CMB) anisotropy	107

9.4	Modern cosmology	107
9.4.1	Inflation	107
9.4.2	Dark energy (The cosmological constant problem)	107
9.4.3	Concordance model [new favorite model]	107
9.4.4	Open questions	107

1 Overview and Review

General relativity is a theory of gravity, replacing Newton's gravity law for heavy masses to give more precise predictions. However, we should keep in mind that general relativity is not yet compatible with quantum mechanics. There are numerous open problems in physics related to reconciling gravity and quantum mechanics.

1.1 Review of Special Relativity

Special relativity studies the kinematics and dynamics in relatively moving inertial reference frames. Most of special relativity and its consequences are encoded in the Lorentz transformations. A classic example of a Lorentz transformation that is often focused on in introductory special relativity is called the “Lorentz boost” in the x -direction. Note that there is nothing special about x or y or z . The x -Lorentz boost is essentially a coordinate transformation, i.e. given coordinates an event A in frame (S) , we can calculate the coordinates of A in frame (S') .

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

where $\beta = v/c$ and the Lorentz factor

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

Since we are studying relativity, it is important to look at “invariants” - quantities that do not change under transformations. In special relativity, one such invariant is the spacetime invariant:

$$\begin{aligned} (\Delta S)^2 &= (c\Delta\tau)^2 \\ &= (c\Delta t)^2 - (\Delta x)^2 - (\Delta y)^2 - (\Delta z)^2 \\ &= (c\Delta t')^2 - (\Delta x')^2 - (\Delta y')^2 - (\Delta z')^2 \\ &= (\Delta S')^2 = (c\Delta\tau')^2. \end{aligned}$$

This can be readily shown. In fact, we have verified this in *Special Relativity: A Quick Guide*. This quantity, roughly speaking, is a measure of **proper** distances and times. If we go to a rest frame of event A , such that $\Delta x' = \Delta y' = \Delta z' = 0$, then we get $\Delta t' = \Delta\tau$, where τ denotes the **proper time**, which is the time measured in the rest frame. This gives $(\Delta S)^2 = (c\Delta\tau)^2$.

In relativity in general and in Minkowski spacetime in particular, we are interested in two types of objects: **scalars** and **vectors**. Scalars are invariant

under general coordinate transformations. An example of a scalar is the space-time invariant. Another scalar, which we probably will not see again in this text, is the metric signature ($\text{sgn}([\eta_{\mu\nu}] = -2)$). Vectors, on the other hand, are a different type of objects, which transform in the same way under coordinate transformations, but are not invariant under general coordinate transformations in general, i.e., their components are not necessarily the same in different coordinate systems. In 4-dimensional spacetime, we work with 4-vectors, which simply means 4-component vectors.

In special relativity, we often talk about position vectors $\mathbf{x} = (ct, x, y, z)^\top = (ct, \vec{x})^\top$ and energy-momentum vectors $\mathbf{p} = (E/c, p_x, p_y, p_z)^\top = (E/c, \vec{p})^\top$. These vectors transform under the Lorentz transformation.

Example 1.1. The x -Lorentz boost applied to \mathbf{p} gives

$$\begin{pmatrix} E'/c \\ p'_x \\ p'_y \\ p'_z \end{pmatrix} = \begin{pmatrix} \gamma & -\gamma\beta & 0 & 0 \\ -\gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} E/c \\ p_x \\ p_y \\ p_z \end{pmatrix}.$$

Notice that E/c transforms like ct (time) and \vec{p} transforms like \vec{x} (space). We also have the following invariant

$$\frac{E^2}{c^2} - p_x^2 - p_y^2 - p_z^2 = \frac{E^2}{c^2} - \vec{p} \cdot \vec{p}.$$

But recall that $(mc)^2 = E^2/c^2 - \vec{p} \cdot \vec{p}$, so

$$\frac{E^2}{c^2} - p_x^2 - p_y^2 - p_z^2 = (mc)^2.$$

If we go to a rest frame, such that $\vec{p} = \vec{0}$, then we obtain the famous rest mass-energy equivalence: $E = mc^2$.

Some quantities associated with a vector can be invariant (scalar), such as their norm. In fact, the spacetime invariant in Minkowski space is nothing but a dot product of a vector with itself:

$$\mathbf{a} \cdot \mathbf{a} = a^0 a^0 - a^1 a^1 - a^2 a^2 - a^3 a^3,$$

where $0, 1, 2, 3$ are indices, and \mathbf{a} implies a 4-vectors, which should be distinguished from the 3-vector $\vec{a} = (a^1, a^2, a^3)^\top$. Notice that the dot product in Minkowski spacetime is defined differently from that in flat 3-dimensional Cartesian coordinate system. In the following chapters, we will explore how dot products are defined in general. Hint: the metric tensor plays an important role.

1.2 The Equivalence Principle

In 1907, Albert Einstein had the “happiest thought of his life” when he realized that in a freely falling frame (non rotating and/or accelerating), the effects of gravity go away, i.e., there is an equivalence between gravity and acceleration such that they can “undo” each other. For example, the following two situations are equivalent in terms of the acceleration experienced by the observer: (i) a person standing on Earth, and (ii) a person inside an elevator accelerating upwards at rate g in free space (no gravitational field). Likewise, the following two situations are also equivalent: (iii) a person floating in free space, and (iv) a person inside an airplane free falling towards Earth (this is commercially known as “zero-G flight”). we arrive at the statement of the Equivalence Principle:

“A small, non-rotating, freely falling frame in a \vec{g} field is an inertial frame.”

The above statement is a direct result of Galileo’s discovery that all objects have the same acceleration due to gravity. This result may seem a little bit *circular*, because it is actually a coincidence that the two roles of **mass**: (i) to cause gravitational force like charge in an electric field:

$$\vec{F} = \frac{GMm_G}{r^3}\vec{r} = m_G\vec{g}$$

and (ii) to measure inertia:

$$\vec{F} = m_I\vec{a}$$

are the same, i.e., $m_I = m_G$. It could have been that “gravitational mass” m_G and “inertial mass” m_I are not the same, in which case $\vec{a} \neq \vec{g}$, and the equivalence principle does not hold. However, the famous Eötvös experiment, which measured the correlation between inertial mass and gravitational mass, has shown that

$$\frac{|m_G - m_I|}{m_I} \leq 10^{-10}.$$

While the observations we have made so far can seem self-apparent, the consequence of the equivalence principle is the bending of light around massive objects. Consider a light ray going horizontally (left to right) pass an upward-accelerating elevator with an observer inside. For the observer, since he is accelerating upwards, he sees the light beam as bent, entering at the top-left of the elevator and exiting at the bottom-right. Now, according to the equivalence principle, because the upward-accelerating scenario is equivalent to the existence of a massive object (such that the observer experiences the same acceleration). This means that the observer, now no longer in the elevator, “sees” the light as bent around this massive object. Fig. 1. illustrates this postulate.

General relativity predicts that light going pass Earth’s surface will “fall” by approximately 1\AA , which is not observable. However, for a much more massive

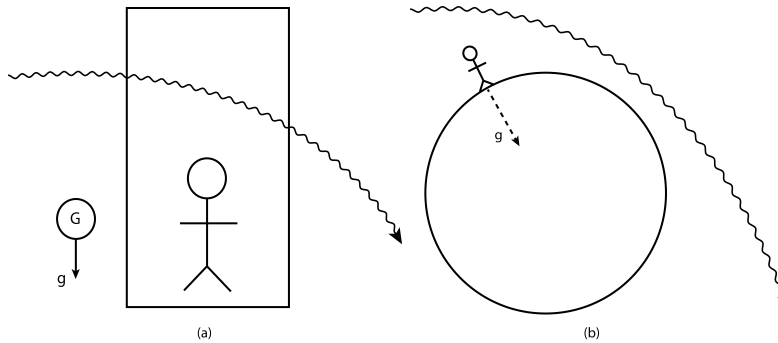


Figure 1: (a) A person in an upward-accelerating elevator (hence “Ground” is downward-accelerating at rate g) sees light as bent. (b) A person on a massive planet also sees light as bent.

object like the Sun, general relativity predicts a bending of $1.75''$ (arc sec). This prediction was verified by the glorious experiment of Arthur Eddington.

Note that we could argue for the bending of light, using Newtonian physics. However, in order to get the correct predictions for the bending of light, we need general relativity. We will explore the reason behind this discrepancy in the following chapter. But roughly speaking, spacetime is assumed to be flat in Newtonian physics, while spacetime is curved by massive objects, according to general relativity. One might ask: “How do we view falling objects on Earth as due to the curvature of spacetime?” The answer requires bringing back Minkowski spacetime diagrams.

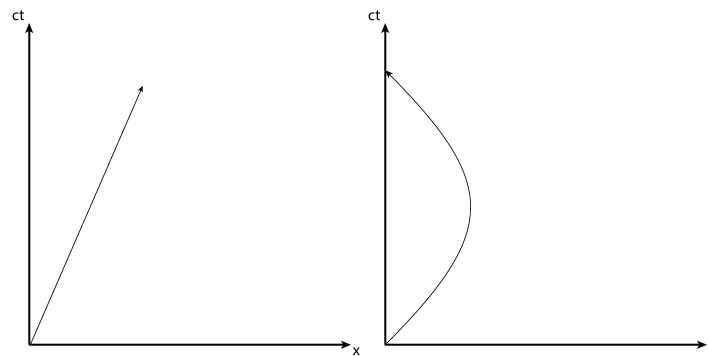


Figure 2: World-lines of an object traveling at constant velocity v and of an object that is falling. Note that the “curvature” due to g here is extremely exaggerated.

1.3 Versions of the Equivalence Principle

There are two versions of the equivalence principle, referred to as the *strong equivalence principle*, SEP, and the *weak equivalence principle*, WEP.

1.3.1 The Strong Equivalence Principle

The strong equivalence principle states that **all** of physics reduces to special relativity in a freely falling frame.

1.3.2 The Weak Equivalence Principle

The weak equivalence principle states that all point particles fall at the same rate in a gravitational field ($m_G = m_I$). We notice that WEP only applies to gravity, which makes it sufficient to develop general relativity, but not quantum mechanics. For this remaining of this note, we will rely on the WEP.

2 Review of Vector Calculus

2.1 Operations & Theorems

Recall that a path in 3-dimensional space can be parameterized by a single variable, which we will call t , as

$$\vec{r}(t) = x(t)\hat{i} + y(t)\hat{j} + z(t)\hat{k},$$

where the tangent to this path is given by

$$\dot{\vec{r}}(t) = \frac{d\vec{r}}{dt}.$$

The length of the underlying curve of this path is the integral over the line element $ds = ||\vec{r}'|| = ||\dot{\vec{r}}|| dt$:

$$L = \int_a^b dS = \int_a^b ||\dot{\vec{r}}|| dt.$$

Example 2.1. Consider a path in \mathbb{R}^2 defined as $\vec{r}(t) = (t, \sin t)$, with $t \in [-2\pi, 2\pi]$. We instantly recognize that the underlying curve of this path is nothing but a sine curve in the xy -plane. The length of this curve is

$$L = \int_{-2\pi}^{2\pi} ||(1, \cos t)|| dt = \int_{-2\pi}^{2\pi} \sqrt{1 + \cos^2 t} dt \approx 15.28.$$

Next, we consider a vector-valued function $\vec{F} : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ defined as $\vec{F}(\vec{r}) = (F_x(x, y, z), F_y(x, y, z), F_z(x, y, z))$. Recall the two “del operators” from vector calculus: the **curl** and the **div**, whose definitions are

$$\text{curl}(\vec{F}) = \vec{\nabla} \times \vec{F} = \det \begin{pmatrix} \hat{i} & \hat{j} & \hat{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ F_x & F_y & F_z \end{pmatrix}$$

and

$$\text{div}(\vec{F}) = \vec{\nabla} \cdot \vec{F} = \frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z}.$$

where the “del” symbol represents the operator

$$\vec{\nabla} = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right).$$

We have also seen the **gradient** of a scalar field (or potential) $f : \mathbb{R}^3 \rightarrow \mathbb{R}$. The gradient is defined as

$$\vec{\nabla} f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z} \right).$$

There are two kinds of potential functions in vector calculus: scalar potential and vector potential. While we are more familiar with the scalar potential (e.g. the gravitational potential or the electric potential in electromagnetism), there isn't as strong a connection between vector potentials and any physical meaning. However, we should still look at some examples.

Example 2.2. In electromagnetism, an electric field \vec{E} is defined as the negative gradient of the electric potential ϕ , measured in Volts:

$$\vec{E} = -\vec{\nabla}\phi.$$

On the other hand, an magnetic field \vec{B} is defined as the curl of the magnetic potential \vec{A} .

$$\vec{B} = \vec{\nabla} \times \vec{A}.$$

In this case, A is a vector potential. Note that we don't often talk about a vector potential of an electric field. This is simply because for a point charge, which produces an inverse-square field, the vector potential simply does not exist. The proof is straightforward. For curious readers, this proof can be an interesting exercise (the solution should be a one-liner). Or else, please refer to my vector calculus notes on my website. Hint: we need one of the two following fundamental theorems of vector calculus: Stokes' and Gauss' (or Ostrogradsky's).

Stokes' theorem states that, the circulation around a closed curved in \mathbb{R}^3 by a vector field \vec{F} is the flux of $\vec{\nabla} \times \vec{F}$ through the surface S whose boundary $C = \partial S$ is the curve C .

$$\oint_{C=\partial S} \vec{F} \cdot d\vec{s} = \iint_S \vec{\nabla} \times \vec{F} \cdot d\vec{S}.$$

Gauss' theorem states that, if S is closed surface, oriented outward, and if \vec{F} is defined throughout the solid region W enclosed by S , then the flux of \vec{F} through S is the total divergence of S through the region W .

$$\oiint_{S=\partial W} \vec{F} \cdot d\vec{S} = \iiint_W \vec{\nabla} \cdot \vec{F} dV.$$

Line integrals are one of a few important operations in vector calculus. In the context of undergraduate or high school physics, we often use the line integral to calculate the work done by a force \vec{F} over some path \vec{r} . The integral in general is simply an accumulation of the component of the vector field along the path.

$$\int_a^b \vec{F} \cdot d\vec{r}.$$

If \vec{F} is a conservative field, then the line integral along \vec{F} between two point A and B only depends on where A and B are, i.e., the line integral is path-independent. A good illustration of this fact is the way we compute the change in electric potential:

$$W = - \int \vec{E} \cdot d\vec{r} = \Delta\phi$$

To actually compute a line integral, we often look for symmetry first (or whether \vec{F} is conservative), then parameterize $\vec{r} = \vec{r}(s)$ if we must. Then, the line integral becomes

$$\int_a^b \vec{F} \cdot d\vec{r} = \int_a^b \vec{F}(\vec{r}(s)) \cdot \frac{d\vec{r}}{ds} ds$$

Surface integrals give the flux of a vector field through a surface:

$$\int \vec{F} \cdot d\vec{S} = \int \vec{F} \cdot \vec{n} dS = \pm \int \vec{F} \cdot \vec{N} d\phi, d\theta,$$

where \vec{F} is the vector field, and \vec{n} is the unit normal vector to the surface. The last expression denotes the surface integral in terms of a parameterization $\vec{S}(\phi, \theta) = (x(\phi, \theta), y(\phi, \theta), z(\phi, \theta))$. \vec{N} is the standard normal to the surface, which is not necessarily a unit vector (with norm of 1) and not necessarily in the same direction as \vec{n} , which is intrinsic to the surface.

Example 2.3. Consider the electric flux through a sphere:

$$\Phi_E = \oiint \vec{E} \cdot d\vec{a} = \frac{q}{\epsilon_0}.$$

The last expression comes from Gauss' (or Ostrogradsky's) theorem, equivalently the divergence theorem, where q is the charge enclosed.

We should introduce the Maxwell's equations as they "bring together" the concepts we have touched on so far. To see how, we consider the Maxwell's equations in integral form:

Gauss' law: $\boxed{\oiint \vec{E} \cdot d\vec{a} = \frac{q}{\epsilon_0}}$

No magnetic monopole: $\boxed{\oint \vec{B} \cdot d\vec{a} = 0}$

Faraday's law: $\boxed{\oint \vec{E} \cdot d\vec{s} = -\frac{d}{dt}\Phi_B = \frac{\partial}{\partial t} \int_A \vec{B} \cdot d\vec{a}}$

Ampere-Maxwell's law: $\boxed{\oint \vec{B} \cdot d\vec{s} = \mu_0 I + \mu_0 \epsilon_0 \frac{\partial}{\partial t} \int_A \vec{E} \cdot d\vec{a}}$

Using the theorems and facts we have discussed, we can convert the above equations into differential form. First, we can use Gauss' theorem on Gauss' law, with

$$q = \int_V \rho d^3r$$

where ρ is the charge volume density. So, Gauss' law becomes

$$\oint \vec{E} \cdot d\vec{a} = \int_V \vec{\nabla} \cdot \vec{E} d^3r = \frac{1}{\epsilon_0} \int_V \rho d^3r,$$

which implies that the divergence of any electric field \vec{E} is proportional to the charge enclosed q .

$$\boxed{\vec{\nabla} \cdot \vec{E} = \frac{\rho}{\epsilon_0}}$$

Applying Gauss' theorem to the second equation,

$$\oint \vec{B} \cdot d\vec{a} = \int_V \vec{\nabla} \cdot \vec{B} d^3r = 0$$

we immediately see that the divergence of any magnetic field \vec{B} is 0, i.e., there is no such thing as a “magnetic monopole.”

$$\boxed{\vec{\nabla} \cdot \vec{B} = 0}$$

We can apply Stokes' theorem on the other two equations to turn them into differential form. Converting the third equation is as simple as bookkeeping:

$$\oint \vec{E} \cdot d\vec{s} = \int_A \vec{\nabla} \times \vec{E} \cdot d\vec{a} = -\frac{\partial}{\partial t} \int_A \vec{B} \cdot d\vec{a},$$

which says the curl of any electric field \vec{E} is negatively proportional to the change in the magnetic field.

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

To convert the fourth equation into differential form, we define a new quantity, \vec{J} , as current (area) density, such that

$$I = \int_A \vec{J} \cdot d\vec{a}.$$

Again, applying Stokes' theorem to the last equation,

$$\oint \vec{B} \cdot d\vec{s} = \int_A \vec{\nabla} \times \vec{B} \cdot d\vec{a} = \mu_0 I + \mu_0 \epsilon_0 \frac{\partial}{\partial t} \int_A \vec{E} \cdot d\vec{a} = \mu_0 \int_A \left(\vec{J} + \epsilon_0 \frac{\partial \vec{E}}{\partial t} \right) \cdot d\vec{a}.$$

So, in differential form:

$$\vec{\nabla} \times \vec{B} = \mu_0 \vec{J} + \mu_0 \epsilon_0 \frac{\partial \vec{E}}{\partial t}$$

In a later chapter, we will see how to make these equations fully relativistic. For the curious reader wanting more information regarding this subsection, please refer to a standard textbook on vector calculus, or feel free to use my Vector Calculus lecture notes for a more formal treatment of this subject.

2.2 Coordinate Systems

There are numerous coordinate systems to describe 3-dimensional space. However, we are most familiar with the Cartesian coordinates (x, y, z) , the Spherical coordinates (r, θ, ϕ) , and the Cylindrical coordinates (ρ, ϕ, z) . The Cartesian coordinate system is the most simple, elegant, and “nice” of the three systems, since the (x, y, z) coordinates can also be coefficients of the vector pointing from the origin to any point in space. We will see that this is not the case in general coordinate systems. One such coordinate system is the **spherical coordinate system** (illustrated in Fig. 3), where

$$\begin{aligned} x &= r \sin \theta \cos \phi \\ y &= r \sin \theta \sin \phi \\ z &= r \cos \theta \end{aligned}$$

and the inverse relations are

$$\begin{aligned} r &= \sqrt{x^2 + y^2 + z^2} \\ \theta &= \cos^{-1} \left(\frac{z}{r} \right) = \cos^{-1} \left(\frac{z}{\sqrt{x^2 + y^2 + z^2}} \right) \\ \phi &= \tan^{-1} \left(\frac{y}{x} \right) \end{aligned}$$

where $0 \leq r$, $0 \leq \theta \leq \pi$, and $0 \leq \phi \leq 2\pi$. The **volume element** is given by

$$dV = dx dy dz = r^2 \sin \theta dr d\theta d\phi.$$

Consider a sphere of radius R . The **area element** is given by

$$dA = R d\theta d\phi.$$

In **cylindrical coordinates**,

$$\begin{aligned} x &= \rho \cos \phi \\ y &= \rho \sin \phi \\ z &= z \end{aligned}$$

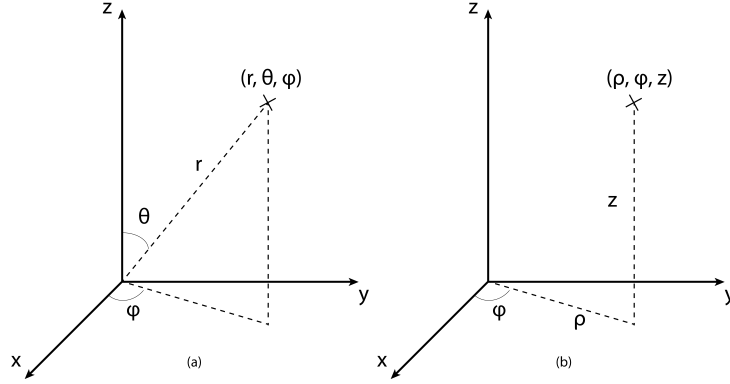


Figure 3: (a) Spherical coordinates, (b) Cylindrical coordinates

and the inversion relations are

$$\begin{aligned}\rho &= \sqrt{x^2 + y^2} \\ z &= z \\ \phi &= \tan^{-1} \left(\frac{y}{x} \right)\end{aligned}$$

The **volume element** is given by

$$dV = r \, dt \, dz \, d\phi.$$

Consider a cylinder of radius R , the **area element** is given by

$$dA = R \, dz \, d\phi.$$

The purpose of having the area and volume elements is so that we could integrate over regions, parameterized in spherical or cylindrical coordinates. While it seems like these quantities appear out of the blue, there is a general procedure to find them, given any arbitrary coordinate system. This “procedure” is called the **Jacobian**, which essentially gives the scaling factor as we move from one coordinate system into another. We know that the area element in polar coordinate is $dA = r \, dr \, d\theta$. We can see this is true, using the Jacobian. First, the Jacobian matrix, which is characteristic of a transformation from coordinate system u^i into $u^{j'}$ is defined as

$$[U_i^{j'}] = \left[\frac{\partial u^{j'}}{\partial u^i} \right] = \begin{pmatrix} \frac{\partial u^1}{\partial u^{1'}} & \cdots & \frac{\partial u^1}{\partial u^{n'}} \\ \vdots & \ddots & \vdots \\ \frac{\partial u^n}{\partial u^{1'}} & \cdots & \frac{\partial u^n}{\partial u^{n'}} \end{pmatrix}$$

Note that i and j' are indices, not powers. Now, the general **scaling factor** is defined by the following equality

$$N = du^{1'} du^{2'} \dots du^{n'} = \left| \det \left([U_i^j] \right) \right| du^1 du^2 \dots du^n$$

Example 2.4. The scaling factor as we move from 2-dimensional Cartesian coordinates into polar coordinates is

$$N = \det \begin{pmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{pmatrix} = \det \begin{pmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{pmatrix} = r.$$

So, the area element is $dA = dx dy = r dr d\theta$, as expected.

Example 2.5. We can try repeating the same procedure to get the volume element for spherical coordinates, which we have said before to be $dV = r^2 \sin \theta dr d\theta d\phi$. First, define the Jacobian matrix:

$$[U_i^{j'}] = \begin{pmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} & \frac{\partial x}{\partial \phi} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} & \frac{\partial y}{\partial \phi} \\ \frac{\partial z}{\partial r} & \frac{\partial z}{\partial \theta} & \frac{\partial z}{\partial \phi} \end{pmatrix} = \begin{pmatrix} \sin \theta \cos \phi & r \cos \theta \cos \phi & -r \sin \theta \sin \phi \\ \sin \theta \sin \phi & r \cos \theta \sin \phi & r \sin \theta \cos \phi \\ \cos \theta & -r \sin \theta & 0 \end{pmatrix}$$

After some bookkeeping-type rearrangements and computation, we find that

$$N = \left| \det [U_i^{j'}] \right| = r^2 \sin \theta,$$

as expected. As a check, we can integrate this volume element over a region defined by $x^2 + y^2 + z^2 = R^2$ to find its volume

$$\int_0^R \int_0^\pi \int_0^{2\pi} r^2 \sin \theta dr d\theta d\phi = \frac{4}{3} \pi R^3,$$

which is not surprisingly the volume of the sphere of radius R .

3 Flat 3-dimensional space

Flat, or “Euclidean” space are spaces with **no curvature**. For instance, a flat sheet of paper in \mathbb{R}^3 is considered “flat” while the surface of a sphere embedded in \mathbb{R}^3 is considered curved space. There should also be a distinction between a sphere in \mathbb{R}^3 and a sphere **embedded** into \mathbb{R}^3 . The former case is simply some subset of flat 3-dimensional space, so it is considered flat. But in the latter case, the sphere is a curved 2-dimensional space, although it lives in flat 3-dimensional space.

Our goal in this section is to see how we can use any arbitrary coordinate system, which specify points in 3-dimensions as an intersection of 3 surfaces. For instance, in Cartesian coordinates, a point (a^1, a^2, a^3) is specified by 3 planes: $x = a^1$, $y = a^2$, and $z = a^3$. In spherical coordinates, a point is specified by intersecting a sphere of radius r , a cone with azimuth angle θ , and a plane of constant ϕ . In cylindrical coordinates, a point is specified by intersecting a cylinder of radius ρ , vertical plane of constant ϕ to the x -axis, and the horizontal plane of constant z .

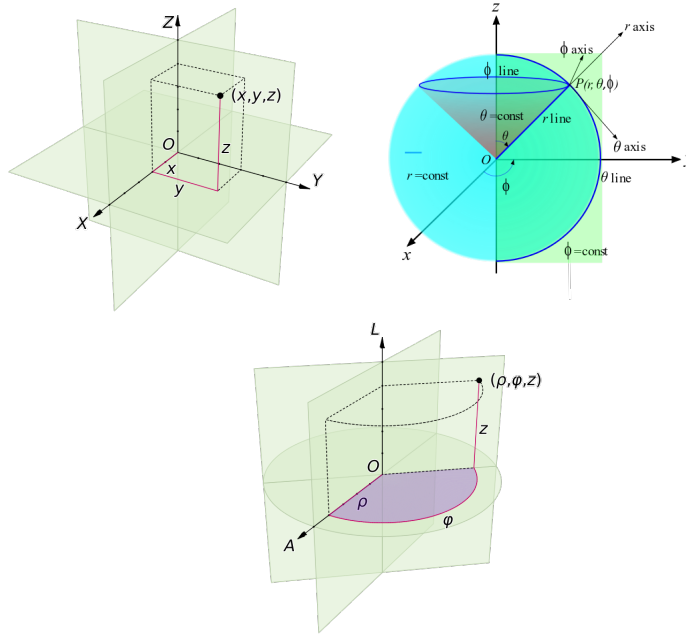


Figure 4: (a) Three intersecting (mutually distinct) planes define a point in Cartesian coordinates ([source](#)), (b) Three intersecting surfaces (a cone, a sphere, and a plane) defining a point in spherical coordinates ([source](#)), and (c) Three intersecting surfaces (two planes and a cylinder) defining a point in cylindrical coordinates ([source](#)).

3.1 Curvilinear coordinates

Curvilinear coordinates are an arbitrary coordinate system in Euclidean space where coordinate lines can be curved (hence curvilinear). Note that although coordinate lines can be curvy, the space is still flat. We can call (u, v, w) our arbitrary coordinates, which specify a point (a^1, a^2, a^3) by intersecting three surfaces: $u = a^1$, $v = a^2$, $w = a^3$. Let relations between (u, v, w) and (x, y, z) be established as

$$u = u(x, y, z); \quad x = x(u, v, w)$$

$$v = v(x, y, z); \quad y = y(u, v, w)$$

$$w = w(x, y, z); \quad z = z(u, v, w)$$

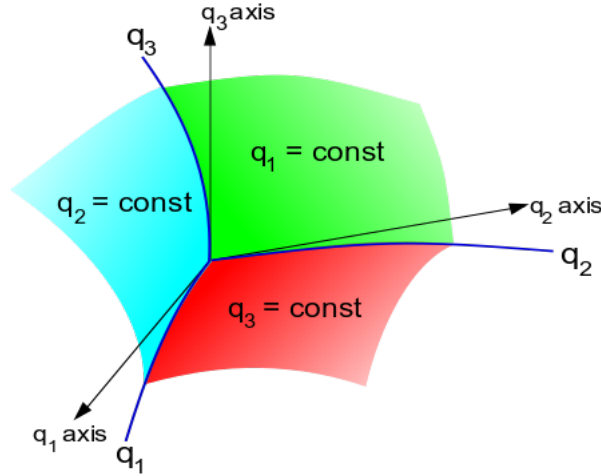


Figure 5: In curvilinear coordinates, a point is defined by the intersection of level surfaces ([source](#)).

3.2 Basis vectors

3.3 Natural basis

Next, since we want to be able to describe vectors using arbitrary curvilinear coordinates, we need to know a **basis set** of vectors that span the space. In Cartesian coordinates, such a basis set is $\{\hat{i}, \hat{j}, \hat{k}\}$. Similarly, we want to find a

set $\{\vec{e}_u, \vec{e}_v, \vec{e}_w\}$ that would give a basis for an arbitrary curvilinear coordinate system. The way we obtain this set is the same as how we obtain $\{\hat{i}, \hat{j}, \hat{k}\}$ from Cartesian coordinates. We notice that \hat{i} is a vector that follows the change in the x -direction with y and z fixed, i.e., \hat{i} is a tangent vector to a vector \vec{r} along the change in x . Hence

$$\hat{i} = \frac{\partial \vec{r}}{\partial x}.$$

Likewise, $\hat{j} = \frac{\partial \vec{r}}{\partial y}$, and $\hat{k} = \frac{\partial \vec{r}}{\partial z}$. By analogy, we can define the set $\{\vec{e}_u, \vec{e}_v, \vec{e}_w\}$ with

$$\vec{e}_u = \frac{\partial \vec{r}}{\partial u}, \quad \vec{e}_v = \frac{\partial \vec{r}}{\partial v}, \quad \vec{e}_w = \frac{\partial \vec{r}}{\partial w},$$

In general, given a coordinate system with (u^1, u^2, \dots, u^n) , we can define a **natural basis set** by letting

$$\boxed{\vec{e}_{u^i} = \frac{\partial \vec{r}}{\partial u^i}}$$

where $i = 0, 1, 2, \dots, n$ are indices. Note that the set $\{\vec{e}_{u^i}\}$ need not be orthogonal. The only requirement is that they have to be linearly independent and span the space. We also note that they need not be unit vectors. It is possible to make a basis set of only unit vectors by letting

$$\hat{e}_u = \frac{\vec{e}_u}{\|\vec{e}_u\|},$$

but we will see that this definition is not so useful in general relativity. A question one might ask is: “What is **natural** about this basis set?” The answer, unfortunately, will be provided in the following sections, where we will discuss how they give rise to the **metric tensor**.

We will often use $\{\hat{i}, \hat{j}, \hat{k}\}$ as a reference basis, i.e., we can express $\{\vec{e}_u, \vec{e}_v, \vec{e}_w\}$ in terms of these:

$$\vec{e}_u = (e_u)_x \hat{i} + (e_u)_y \hat{j} + (e_u)_z \hat{k}.$$

Example 3.1. We can look at the basis set for the spherical coordinate system, $\{\vec{e}_r, \vec{e}_\theta, \vec{e}_\phi\}$. First, let the coordinates (u, v, w) be (r, θ, ϕ) , the spherical coordinates. A vector \vec{r} , expressed in spherical coordinates, has the form:

$$\vec{r} = \begin{pmatrix} r \sin \theta \cos \phi \\ r \sin \theta \sin \phi \\ r \cos \theta \end{pmatrix},$$

where $r \geq 0$, $0 \leq \theta \leq \pi$, and $0 \leq \phi \leq 2\pi$. To obtain the basis vectors, we simply take $\partial \vec{r} / \partial r$, $\partial \vec{r} / \partial \theta$, and $\partial \vec{r} / \partial \phi$. We expect that our resulting vectors

should form the column vectors of the Jacobian matrix $[U_i^{j'}]$, which, as we have discussed previously, has the following form:

$$[U_i^{j'}] = \begin{pmatrix} \sin \theta \cos \phi & r \cos \theta \cos \phi & -r \sin \theta \sin \phi \\ \sin \theta \sin \phi & r \cos \theta \sin \phi & r \sin \theta \cos \phi \\ \cos \theta & -r \sin \theta & 0 \end{pmatrix}.$$

What we're doing now, in terms of linear algebra, is essentially constructing $[U_i^{j'}]$ - the change-of-basis matrix. Let's see if our described procedure works:

$$\begin{aligned} \vec{e}_r &= \frac{\partial \vec{r}}{\partial r} = \begin{pmatrix} \sin \theta \cos \phi \\ \sin \theta \sin \phi \\ \cos \theta \end{pmatrix} \\ \vec{e}_\theta &= \frac{\partial \vec{r}}{\partial \theta} = \begin{pmatrix} r \cos \theta \cos \phi \\ r \cos \theta \sin \phi \\ -r \sin \theta \end{pmatrix} \\ \vec{e}_\phi &= \frac{\partial \vec{r}}{\partial \phi} = \begin{pmatrix} -r \sin \theta \sin \phi \\ r \sin \theta \cos \phi \\ 0 \end{pmatrix} \end{aligned}$$

Not surprisingly,

$$[U_i^{j'}] = \begin{pmatrix} \sin \theta \cos \phi & r \cos \theta \cos \phi & -r \sin \theta \sin \phi \\ \sin \theta \sin \phi & r \cos \theta \sin \phi & r \sin \theta \cos \phi \\ \cos \theta & -r \sin \theta & 0 \end{pmatrix} = (\vec{e}_r | \vec{e}_\theta | \vec{e}_\phi).$$

Next, we can check if $\vec{e}_r, \vec{e}_\theta, \vec{e}_\phi$ is mutually-perpendicular set. (Note that I avoid the word "orthogonal" here, since it carries different meanings in physics and in linear algebra). We can also check whether the three vectors are unit vectors. It suffices to take the dot products of $\vec{e}_r, \vec{e}_\theta, \vec{e}_\phi$ among themselves:

$$\begin{aligned} \vec{e}_r \cdot \vec{e}_r &= \sin^2 \theta (\cos^2 \phi + \sin^2 \phi) + \cos^2 \theta = 1 \\ \vec{e}_r \cdot \vec{e}_\theta &= r \cos^2 \phi \sin \theta \cos \theta + r \sin^2 \phi \sin \theta \cos \theta - r \sin \theta \cos \theta = 0 \\ \vec{e}_r \cdot \vec{e}_\phi &= -r \sin^2 \theta \sin \phi \cos \phi + r \sin^2 \theta \sin \phi \cos \phi = 0 \\ \vec{e}_\theta \cdot \vec{e}_\theta &= r^2 \cos^2 \phi \cos^2 \theta + r^2 \cos^2 \phi \sin^2 \theta + r^2 \sin^2 \theta = r^2 \\ \vec{e}_\theta \cdot \vec{e}_\phi &= 0 \\ \vec{e}_\phi \cdot \vec{e}_\phi &= r^2 \sin^2 \phi \sin^2 \theta + r^2 \sin^2 \theta \cos^2 \phi = r^2 \sin^2 \theta. \end{aligned}$$

We notice that all dot products of cross-coordinates result in 0, which means the basis vectors we found are mutually perpendicular (we can also see this fat geometrically, based on how the spherical coordinates are defined). But we also find that these basis vectors are not unit vectors, since their norms are not necessarily 1:

$$\begin{aligned} \|\vec{e}_r\| &= 1 \\ \|\vec{e}_\theta\| &= r \\ \|\vec{e}_\phi\| &= r \sin \theta. \end{aligned}$$

3.4 Dual basis

The procedure we have described in the previous subsection allows us to find a **natural basis** set of vectors for a coordinate system. However, basis sets are not unique. In this section, we look at an alternative basis set of vectors, called the **dual basis** set, denoted $\{\vec{e}^u, \vec{e}^v, \vec{e}^w, \dots\}$. Instead of deriving these from the finding the tangent vectors, we find the normal of surfaces of constant u, v, w, \dots

Recall that the **gradient** operator $\vec{\nabla}$, when applied to a scalar field f , gives the normal vectors to surfaces of constant f . Since curvilinear coordinates are given by coordinate lines where u, v , or w is constant, we can let

$$\begin{aligned}\vec{e}^u &= \vec{\nabla}u \\ \vec{e}^v &= \vec{\nabla}v \\ \vec{e}^w &= \vec{\nabla}w,\end{aligned}$$

and so on, where \vec{e}^u is a normal vector to the surface of constant u . Or more generally,

$$\boxed{\vec{e}^{u^i} = \vec{\nabla}u^i}$$

Now, we might wonder what the dual basis in Cartesian coordinates is. It turns out the dual and natural basis sets in Cartesian coordinates are the same:

$$\begin{aligned}\vec{e}^x &= \vec{\nabla}x = (1, 0, 0)^\top = \hat{i} = \vec{e}_x \\ \vec{e}^y &= \vec{\nabla}y = (0, 1, 0)^\top = \hat{j} = \vec{e}_y \\ \vec{e}^z &= \vec{\nabla}z = (0, 0, 1)^\top = \hat{k} = \vec{e}_z.\end{aligned}$$

There's actually no magic behind this "coincidence." The reason the basis sets are the same is simply that the direction of increasing/decreasing x is the same as the normal vector to the surface of constant x , i.e., the yz -plane.

insert figure 6 here!!!

However, we suspect that this is not the case in general curvilinear coordinates, since the direction of increasing u , for instance, should not necessarily be the same as the normal vector to the surface of constant u , as illustrated here:

insert figure 7 here!!!

Example 3.2. We can look at the dual basis set for the spherical coordinate system. Based on what we have established so far, we can compute the dual basis set, using the inverse relations $(x, y, z) \rightarrow (r, \theta, \phi)$.

$$\vec{e}^r = \vec{\nabla}r = \vec{\nabla}(x^2 + y^2 + z^2)^{1/2} = (x^2 + y^2 + z^2)^{-1/2} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} \sin \theta \cos \phi \\ \sin \theta \sin \phi \\ \cos \theta \end{pmatrix}.$$

Tedious and painstaking computations give us the other two vectors:

$$\begin{aligned}\vec{e}^\theta &= \frac{1}{r} \begin{pmatrix} \cos \theta \cos \phi \\ \cos \theta \sin \phi \\ -\sin \theta \end{pmatrix} \\ \vec{e}^\phi &= \frac{1}{r} \begin{pmatrix} -\frac{\sin \phi}{\sin \theta} \\ \frac{\cos \phi}{\sin \theta} \\ 0 \end{pmatrix}\end{aligned}$$

We can readily compare $\{\vec{e}^r, \vec{e}^\theta, \vec{e}^\phi\}$ to $\{\vec{e}_r, \vec{e}_\theta, \vec{e}_\phi\}$ and find that while $\vec{e}^r = \vec{e}_r$, $\vec{e}^\theta \neq \vec{e}_\theta$ and $\vec{e}^\phi \neq \vec{e}_\phi$.

Example 3.3. Let's consider a non-familiar case of paraboloidal surfaces (u, v, w) , a non-orthogonal set, where

$$\begin{aligned}x &= u + v \\ y &= u - v \\ z &= 2uv + w,\end{aligned}$$

which gives the inverse relations

$$\begin{aligned}u &= \frac{1}{2}(x + y) \\ v &= \frac{1}{2}(x - y) \\ w &= z - \frac{1}{2}(x^2 - y^2).\end{aligned}$$

Let $\vec{r} = (x, y, z)^\top = (u + v, u - v, 2uv + w)^\top$. We first obtain the natural basis set:

$$\begin{aligned}\vec{e}_u &= \frac{\partial \vec{r}}{\partial u} = (1, 1, 2v)^\top \\ \vec{e}_v &= \frac{\partial \vec{r}}{\partial v} = (1, 1, 2u)^\top \\ \vec{e}_w &= \frac{\partial \vec{r}}{\partial w} = (0, 0, 1)^\top,\end{aligned}$$

which is an mutually perpendicular set, by inspection (i.e., all cross-dot product terms are zero). Next, we obtain the dual basis set:

$$\begin{aligned}\vec{e}^u &= \vec{\nabla}u = \frac{1}{2}\vec{\nabla}(x + y) = \left(\frac{1}{2}, \frac{1}{2}, 0\right)^\top \\ \vec{e}^v &= \vec{\nabla}v = \frac{1}{2}\vec{\nabla}(x - y) = \left(\frac{1}{2}, -\frac{1}{2}, 0\right)^\top \\ \vec{e}^w &= \vec{\nabla}w = \vec{\nabla}\left(z - \frac{1}{2}(x^2 - y^2)\right) = (-x, y, 1)^\top = (-u - v, u - v, 1)^\top\end{aligned}$$

But notice that we get non-zero cross-dot product terms:

$$\begin{aligned}\vec{e}^u \cdot \vec{e}^w &= -v \\ \vec{e}^u \cdot \vec{e}^v &= 0 \\ \vec{e}^v \cdot \vec{e}^w &= -u.\end{aligned}$$

The reason for observing what value the dot products between basis vectors are will become clear when we talk about the **metric tensor** in the next subsections.

3.5 Contravariant and covariant vectors

3.5.1 The suffix notation (vector calculus)

In general coordinate systems, there can be many more than three coordinates, and it can be burdensome to keep track of different letters representing these coordinates, especially if they don't follow a convention. It is convenient to change our notation from "letter-based" to "index-based." For example, instead of using (u, v, w) as coordinates, we will now use (u^1, u^2, u^3) , for $i = 1, 2, 3$. Note that we are using upper indices here. As we will discuss soon, upper and lower indices have different meanings.

Similarly, our notation for basis vectors also change. For the natural basis, we "code" $\{\vec{e}_u, \vec{e}_v, \vec{e}_w\}$ as $\{\vec{e}_i\}$, where $i = 1, 2, 3$. For the dual basis, we "code" $\{\vec{e}^u, \vec{e}^v, \vec{e}^w\}$ as $\{\vec{e}^i\}$ for $i = 1, 2, 3$. Our vector notation also changes as a result, but there are additional rules. Let vector $\vec{\lambda}$ be given. In the **natural basis**, the components of λ is written with **upper** indices,

$$\vec{\lambda} = \lambda^1 \vec{e}_1 + \lambda^2 \vec{e}_2 + \lambda^3 \vec{e}_3 = \sum_{i=1}^3 \lambda^i \vec{e}_i.$$

and in the **dual basis**, we use **lower** indices for the components:

$$\vec{\lambda} = \lambda_1 \vec{e}^1 + \lambda_2 \vec{e}^2 + \lambda_3 \vec{e}^3 = \sum_{i=1}^3 \lambda_i \vec{e}^i$$

3.5.2 The Einstein summation notation

We shall make another "leap in notation" as we introduce the Einstein summation convention:

Any index that appears once up and once down is automatically summed

For example, from now on, we will write $\vec{\lambda}$ as:

$$\vec{\lambda} = \sum_{i=1}^3 \lambda^i \vec{e}_i = \lambda^i \vec{e}_i$$

Note that this summation convention removes us from indicating the minimum and maximum values for our indices, which allows us to keep the mathematics as applicable to any n-dimensional system as possible. We should also notice that indices can be exchangeable, especially in summed situations. For example,

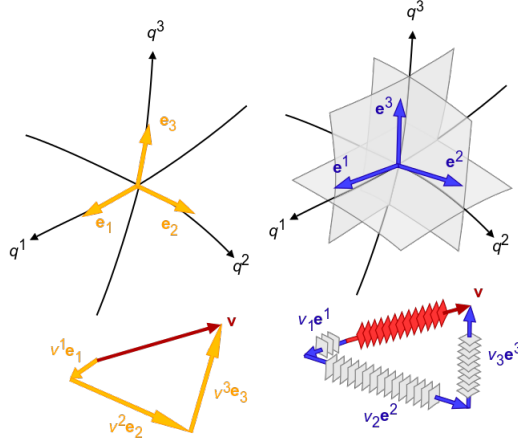
$$a^i b_i = a^j b_j = \dots = \sum_{n=1}^N a^n b_n$$

Keep in mind that $a_i b_i$ does not make sense in the Einstein convention. If we want to denote a sum, we have to specify the range of i as follows:

$$\sum_{i=n}^N a_i b_i.$$

Likewise, $a_i b^i c^i$ is not defined either. Remember that only “1 up index, 1 down index” is allowed. Finally, it is widely agreed upon that certain letters are reserved for some specific cases. For instance, $i, j, k, l, \dots = 1, 2, 3$ are reserved for 3-dimensional space. We use Greek characters $\mu, \nu, \alpha, \beta, \dots = 0, 1, 2, 3$ for 4-dimensional spacetime, $A, B, C, \dots = 1, 2$ for 2-dimensional spaces, and $a, b, c, \dots = 1, 2, \dots, N$ for N-dimensional manifolds.

3.5.3 Contravariant and covariant vectors



Source

Earlier, we have established that any vector $\vec{\lambda}$ can be written in two “upper-lower index” ways, depending on whether we are using the natural or the dual basis set:

$$\vec{\lambda} = \lambda^i \vec{e}_i = \lambda_i \vec{e}^i.$$

We shall now call λ^i the **contravariant** components and λ_i the **covariant** components. (Remember: “co” is “low”). Contravariant vectors are simply vectors with contravariant components. Likewise, covariant vectors are vectors with covariant components. Later in this text, we will denote $\vec{\lambda}$ simply as λ^i (contravariant vector) or λ_i (covariant vector).

Putting together what we have established so far about index notation and Einstein summation convention, we can now look at the last interesting dot product: between a vector in the natural basis and one in the dual basis: $\vec{e}^i \cdot \vec{e}_j$. Before we proceed, keep in mind that this single expression ($\vec{e}^i \cdot \vec{e}_j$) represents 9 different objects because there is not summation here (i is not necessarily j and $i, j = 1, 2, 3$). Now, recall the definitions of these vectors:

$$\begin{aligned}\vec{e}^i &= \vec{\nabla} u^i = \frac{\partial u^i}{\partial x} \hat{i} + \frac{\partial u^i}{\partial y} \hat{j} + \frac{\partial u^i}{\partial z} \hat{k} \\ \vec{e}_j &= \frac{\partial \vec{r}}{\partial u^j} = \frac{\partial x}{\partial u^j} \hat{i} + \frac{\partial y}{\partial u^j} \hat{j} + \frac{\partial z}{\partial u^j} \hat{k}\end{aligned}$$

Therefore,

$$\vec{e}^i \cdot \vec{e}_j = \frac{\partial u^i}{\partial x} \frac{\partial x}{\partial u^j} + \frac{\partial u^i}{\partial y} \frac{\partial y}{\partial u^j} + \frac{\partial u^i}{\partial z} \frac{\partial z}{\partial u^j}$$

But notice that this is nothing but the chain rule, if we think about u^i as $u^i(x, y, z)$. So,

$$\boxed{\vec{e}^i \cdot \vec{e}_j = \frac{\partial u^i}{\partial u^j}}$$

But also notice that the u^i 's are independent variables, i.e.,

$$\frac{\partial u^i}{\partial u^j} = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$$

We can introduce a two-index object, the **Kronecker delta**, defined as

$$\boxed{\delta_j^i = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}}$$

We obtain an elegant relation:

$$\boxed{\vec{e}^i \cdot \vec{e}_j = \delta_j^i}$$

Again, keep in mind that this is not a single equation but rather 9. Six of which are 0 = 0 and three 1 = 1. We observe that for $u \neq v$, $\vec{e}^u \cdot \vec{e}_v = 0$, i.e., $\vec{e}^u \perp \vec{e}_v$, which makes sense, by definition: As a little aside, every time we see

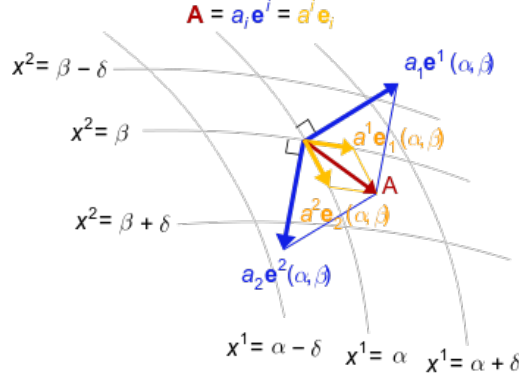


Figure 6: The natural and basis vectors are defined such that $\vec{e}^u \cdot \vec{e}_v = \delta_v^u$. This is not so difficult to show mathematically (all simply follow from the definitions), but a figure can help visualize “what’s going on” ([source](#)).

an two-index object, we should immediately think of a matrix. In matrix form, the Kronecker delta is our beloved identity matrix I :

$$[\delta_j^i] = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

What about the dot (or inner) products of $\{\vec{e}^i\}$ and $\{\vec{e}_j\}$ among themselves? We have calculated the terms before, when we introduced spherical coordinates, but we have not thought about their significance. Let us define another two-index object, called the **metric tensor** by

$$g_{ij} = \vec{e}_i \cdot \vec{e}_j$$

which in matrix form is:

$$[g_{ij}] = \begin{pmatrix} \vec{e}_1 \cdot \vec{e}_1 & \vec{e}_1 \cdot \vec{e}_2 & \vec{e}_1 \cdot \vec{e}_3 \\ \vec{e}_2 \cdot \vec{e}_1 & \vec{e}_2 \cdot \vec{e}_2 & \vec{e}_2 \cdot \vec{e}_3 \\ \vec{e}_3 \cdot \vec{e}_1 & \vec{e}_3 \cdot \vec{e}_2 & \vec{e}_3 \cdot \vec{e}_3 \end{pmatrix}.$$

And likewise, we define the **inverse metric tensor** as

$$g^{ij} = \vec{e}^i \cdot \vec{e}^j$$

which in matrix form is:

$$[g^{ij}] = \begin{pmatrix} \vec{e}^1 \cdot \vec{e}^1 & \vec{e}^1 \cdot \vec{e}^2 & \vec{e}^1 \cdot \vec{e}^3 \\ \vec{e}^2 \cdot \vec{e}^1 & \vec{e}^2 \cdot \vec{e}^2 & \vec{e}^2 \cdot \vec{e}^3 \\ \vec{e}^3 \cdot \vec{e}^1 & \vec{e}^3 \cdot \vec{e}^2 & \vec{e}^3 \cdot \vec{e}^3 \end{pmatrix}.$$

Since dot products are commutative, i.e., $\vec{e}^i \cdot \vec{e}^j = \vec{e}^j \cdot \vec{e}^i$ and $\vec{e}_i \cdot \vec{e}_j = \vec{e}_j \cdot \vec{e}_i$, we get an interesting property:

$$g_{ij} = g_{ji} \text{ and } g^{ij} = g^{ji}$$

In other words, the matrices $[g_{ij}]$ and $[g^{ij}]$ are **symmetric**. Just for sanity check, we can look at $[g_{ij}]$ for Cartesian coordinates, which (not surprisingly), turns out to be the identity matrix, i.e., $[g_{ij}] = [g^{ij}] = [\delta_j^i] = I$, which is symmetric. However, $[g_{ij}]$ are not **diagonal**, in general. A counter-example is $[g_{ij}]$ in paraboloidal coordinates.

We are of course not limited to looking only at dot products of the basis vectors among themselves. Consider the dot product of two vectors $\vec{\mu} = \mu^i \vec{e}_i = \mu_i \vec{e}^i$ and $\vec{\lambda} = \lambda^i \vec{e}_i = \lambda_i \vec{e}^i$. We immediately see that there are more than one way to express $\vec{\mu} \cdot \vec{\lambda}$, all of which are equivalent. Now, we have to be careful with writing the dot product with indices, because $\lambda^i \vec{e}^i \mu^i \cdot \vec{e}^i$ is not defined in the Einstein summation convention. We have to realize that the indices of $\vec{\mu}$ and $\vec{\lambda}$ should be independent, so it suffices to replace i with j as indices for one of the vectors. We are now safe to proceed.

$$\vec{\lambda} \cdot \vec{\mu} = \lambda^i \vec{e}_i \cdot \mu^j \vec{e}_j = \lambda_i \vec{e}^i \cdot \mu_j \vec{e}^j = \lambda_i \vec{e}^i \cdot \mu^j \vec{e}_j = \lambda^i \vec{e}_i \cdot \mu_j \vec{e}^j$$

We have shown that $\vec{e}^i \cdot \vec{e}_j \delta_j^i$, $\vec{e}^i \cdot \vec{e}^j = g^{ij}$, and $\vec{e}_i \cdot \vec{e}_j = g_{ij}$, we obtain a number of relations:

$$\begin{aligned} \vec{\lambda} \cdot \vec{\mu} &= \lambda^i \vec{e}_i \cdot \mu^j \vec{e}_j = g_{ij} \lambda^i \mu^j \\ &= \lambda_i \vec{e}^i \cdot \mu_j \vec{e}^j = g^{ij} \lambda_i \mu_j \\ &= \lambda_i \vec{e}^i \cdot \mu^j \vec{e}_j = \delta_j^i \lambda_i \mu^j = \lambda_i \mu^i = \lambda_j \mu^j \\ &= \lambda^i \vec{e}_i \cdot \mu_j \vec{e}^j = \delta_i^j \lambda^i \mu_j = \lambda^i \mu_i = \lambda^j \mu_j \end{aligned}$$

We can summarize the statements above into five equivalent statements:

$$\vec{\lambda} \cdot \vec{\mu} = g_{ij} \lambda^i \mu^j = g^{ij} \lambda_i \mu_j = \lambda_i \mu^i = \lambda^i \mu_i$$

From the second and fourth expressions, we get

$$g_{ij} \mu^j = \mu_i$$

Likewise, if we look at the third and fifth expressions:

$$g^{ij} \lambda_i = \lambda^j$$

This means we can use the metric tensor g_{ij} and the inverse metric tensor g^{ij} to go back and forth between covariant and contravariant vector components. Or, in a more “bookkeeping” sense, g^{ij} raises an index, while g_{ij} lowers an index. These

operations are useful to remember, since they will become especially handy when we have objects with more than one upper and lower indices.

But we can extract even more information from the above two equalities by combining them:

$$\mu^i = g^{ij} \mu_j = g^{ij} (g_{jk} \mu^k).$$

Note that we need an extra index k here since the indices of the expansion of μ_j are independent of the indices of g^{ij} , only except for j . Now, since it is also true that $\mu^i = \delta_j^i \mu^j$, the following has to be true as well:

$$\boxed{g^{ij} g_{jk} = \delta_k^i}$$

We obtain a similar equality if we start with the covariant component μ_i :

$$\boxed{g_{ij} g^{jk} = \delta_i^k}$$

The above two equalities imply that the matrices $[g^{ij}]$ and $[g_{ij}]$ are **inverses** of each other. Hence, we call g_{ij} the **metric tensor** and g^{ij} the **inverse metric tensor**.

Example 3.4. Consider $\vec{a} = \vec{e}_\theta$ in spherical coordinates. Express \vec{a} with contravariant components, covariant components, and find the norm of \vec{a} , with and without the metric g_{ij} .

Since $\vec{a} = \vec{e}_\theta$, a single natural basis vector, we can easily write \vec{a} as a contravariant vector as

$$a^i = (a^1, a^2, a^3)^\top = (0, 1, 0)^\top.$$

We can calculate the covariant components of \vec{a} using the metric tensor g_{ij} :

$$\begin{aligned} a_1 &= g_{1j} a^j = g_{11} a^1 = 0 \\ a_2 &= g_{2j} a^j = g_{22} a^2 = r^2 \\ a_3 &= g_{3j} a^j = g_{33} a^3 = 0. \end{aligned}$$

So, \vec{a} , with covariant components, is

$$a_i = (0, r^2, 0)^\top$$

Next, we can calculate $\|\vec{a}\|$ in two ways, and both will give the same answer:

$$\begin{aligned} \|\vec{a}\|^2 &= a^i a_i = a^1 a_1 + a^2 a_2 + a^3 a_3 = r^2 \\ \|\vec{a}\|^2 &= g_{ij} a^i a^j = g_{11} a^1 a^1 + g_{22} a^2 a^2 + g_{33} a^3 a^3 = r^2. \end{aligned}$$

So, $\|\vec{a}\| = r$.

3.6 Metric tensor

In the previous subsection we have defined the metric tensor and the inverse metric tensor. In this subsection, we will discuss their physical significance and related properties.

To start, the metric tensor contains information about physical lengths and the geometry of the space. To illustrate this point, let us consider a path $\vec{r} = \vec{r}(t)$ in flat 3-dimensional space (Euclidean). For vector calculus, we know that the length of the underlying curve, denoted as L , is given by

$$L = \int_a^b \|\dot{\vec{r}}\| dt.$$

Let this curve be expressed in a general curvilinear coordinate system with coordinates (u, v, w) , i.e., $\vec{r}(t) = (u(t), v(t), w(t))$, it follows that

$$\begin{aligned} \dot{\vec{r}} &= \frac{\partial \vec{r}}{\partial u} \frac{du}{dt} + \frac{\partial \vec{r}}{\partial v} \frac{dv}{dt} + \frac{\partial \vec{r}}{\partial w} \frac{dw}{dt} \\ &= \vec{e}_u \frac{du}{dt} + \vec{e}_v \frac{dv}{dt} + \vec{e}_w \frac{dw}{dt} \\ &= \vec{e}_i \frac{du^i}{dt}, \end{aligned}$$

in Einstein's notation. So,

$$\|\dot{\vec{r}}\| = \sqrt{\dot{\vec{r}} \cdot \dot{\vec{r}}} = \sqrt{\vec{e}_i \frac{du^i}{dt} \cdot \vec{e}_j \frac{du^j}{dt}} = \sqrt{g_{ij} \frac{du^i}{dt} \frac{du^j}{dt}}$$

Putting everything (so far) together, we get

$$L = \int_a^b \left\| \frac{d\vec{r}}{dt} \right\| dt = \int_a^b \sqrt{g_{ij} \frac{du^i}{dt} \frac{du^j}{dt}} dt$$

Let ds the infinitesimal distance, we get

$$ds = \sqrt{g_{ij} \frac{du^i}{dt} \frac{du^j}{dt}} dt.$$

Bringing the dt into the square root, and squaring both sides, we get

$$ds^2 = g_{ij} du^i du^j$$

Even though ds^2 has units of length squared, we still shall refer to this quantity as the **line element**. Let us look at a few examples to see how the above equality manifests in geometries we are familiar with.

Example 3.5. In Cartesian coordinates, the natural basis vectors are $\{\vec{e}_i\}\{\hat{i}, \hat{j}, \hat{k}\}$. Observe that $g_{ij} = \vec{e}_i \cdot \vec{e}_j = \delta_{ij}$, since the \vec{e}_i 's are both unit vectors and mutually perpendicular. So, as a matrix:

$$[g_{ij}] = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

The line element, $ds^2 = g_{ij} du^i du^j$, is a sum of nine terms. However, we notice that $g_{ij} = 0$ if $i \neq j$ and 1 otherwise. Therefore the line element is:

$$ds^2 = g_{ij} du^i du^j = dx^2 + dy^2 + dz^2,$$

which we are very familiar with. Indeed, the moral of this example is to show that while the form of the line element can be derived geometrically, a rigorous way to obtain the line element is actually through the metric. For Cartesian coordinates, deriving the line element is more or less trivial, because we realize that the above equality is simply Pythagorean theorem in the three dimensions. However, as we will see in the next two examples, it is more difficult to obtain the line element by “inspection,” and that realizing the line element from the metric is more reliable.

Example 3.6. We have shown that, in spherical coordinates:

$$\begin{aligned} \vec{e}_r \cdot \vec{e}_r &= 1 \\ \vec{e}_\theta \cdot \vec{e}_\theta &= r^2 \\ \vec{e}_\phi \cdot \vec{e}_\phi &= r^2 \sin^2 \theta \\ \vec{e}_r \cdot \vec{e}_\theta &= \vec{e}_r \cdot \vec{e}_\phi = \vec{e}_\theta \cdot \vec{e}_\phi = 0, \end{aligned}$$

which give the metric, in matrix form, as:

$$[g_{ij}] = \begin{pmatrix} 1 & 0 & 0 \\ 0 & r^2 & 0 \\ 0 & 0 & r^2 \sin^2 \theta \end{pmatrix}.$$

Therefore, the line element in spherical coordinates is:

$$ds^2 = g_{ij} du^i du^j = dr^2 + r^2 d\theta^2 + r^2 \sin^2 \theta d\phi^2.$$

While one can certainly derive this expression geometrically, it is no doubt more challenging to do than in Cartesian coordinates.

Example 3.7. Find the length of a curve in spherical coordinates with the parameterization $\vec{r}(t) = (r(t), \theta(t), \phi(t)) = (t, \pi/4, 4t)$, where $0 \leq t \leq \pi$.

It is possible to proceed without knowing what the underlying curve of this path is. However, it certainly does not harm to know what length we are trying to find here. First, notice that θ is fixed, so we are limited to a cone, oriented

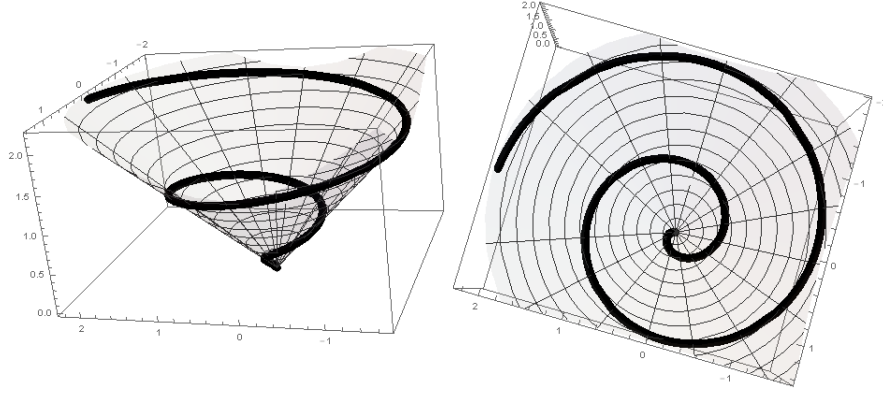


Figure 7: The underlying curve of path $\vec{r}(t)$, for $0 \leq t \leq \pi$, side view and top view.

vertically along the z -axis. Now, since $0 \leq t \leq \pi$, and that $\phi = 4t$, we know that the curve makes two revolutions around the z -axis. So we can imagine the curve looking something like this: Find the length of a curve is only “mechanical” once we already know the line element ds^2 , which we have derived in the previous example. Plus, we know that $d\theta = 0$ since θ is constant. So,

$$\begin{aligned}
 L &= \int ds \\
 &= \int_0^\pi \sqrt{\frac{dr^2}{dt^2} + r^2 \frac{d\theta^2}{dt^2} + r^2 \sin^2 \theta \frac{d\phi^2}{dt^2}} dt \\
 &= \int_0^\pi \sqrt{1 + 0 + \left(4t \sin\left(\frac{\pi}{4}\right)\right)^2} dt \\
 &= \int_0^\pi \sqrt{1 + 8t^2} dt \\
 &\approx 14.55.
 \end{aligned}$$

As we have seen before in the earlier sections, the metric tensor does not just encode physical lengths. g_{ij} are also ubiquitous in calculating norms of vectors and inner products (or dot products). As a reminder:

$$\textbf{Norm: } \|\vec{\lambda}\|^2 = \vec{\lambda} \cdot \vec{\lambda} = g_{ij} \lambda^i \lambda^j$$

$$\textbf{Inner product: } \vec{\lambda} \cdot \vec{\mu} = g_{ij} \lambda^i \mu^j.$$

They can also be used to raise/lower indices, as we have seen.

3.6.1 Summations as matrix products

As a little “aside,” writing vectors and 2-component tensors (such as g_{ij}) as matrices can be more convenient for us. But note that more general tensors

(with three or more indices) can no longer be written as matrices.

Suppose that we have two square matrices $\tilde{A} = [a_{ij}]$ and $\tilde{B} = [b_{ij}]$. Let matrix $\tilde{C} = \tilde{A}\tilde{B}$. In matrix multiplication,

$$\tilde{C} = \begin{pmatrix} \vdots & c_{ij} & \vdots \\ \dots & & \dots \end{pmatrix} = \begin{pmatrix} \vdots & & \vdots \\ a_{i1} & \dots & a_{in} \\ \vdots & & \vdots \end{pmatrix} \begin{pmatrix} \dots & b_{1j} & \dots \\ \vdots & & \vdots \\ \dots & b_{nj} & \dots \end{pmatrix}$$

i.e., the ij^{th} element of \tilde{C} is

$$c_{ij} = \sum_{k=1}^n a_{ik} b_{kj}$$

where k on the a term indicates a column, while k on the b term indicates a row. Notice that the summed index (k) goes as “column-row,” whereas $i-j$ go as row-column. Note that I do not have a restriction on n , because the above relations are true in any n -dimensional space.

Similarly, we can multiply vectors in matrix notations. Let two vectors $\vec{F}, \vec{G} \in \mathbb{R}^n$ be given. Assuming that the metric is the identity, their inner product is given by:

$$\vec{F} \cdot \vec{G} = \begin{pmatrix} f_1 \\ \vdots \\ f_n \end{pmatrix} \cdot \begin{pmatrix} g_1 \\ \vdots \\ g_n \end{pmatrix} = \vec{F}^\top \vec{G} = \sum_{k=1}^n f_k g_k.$$

With a general metric g_{ij} , we want to express $\vec{\lambda} \cdot \vec{\mu} = g_{ij} \lambda^i \mu^j$ as a matrix multiplication. We have to make sure that the matrix multiplication is defined, i.e., the rows of the right-multiplying element is equal to the columns of the left-multiplying element. There are two things we can do: ordering the ij indices, and transposing. Let $\tilde{\lambda}$ denote $\vec{\lambda} = [\lambda^i]$, and $\tilde{\mu}$ denote $\vec{\mu} = [\mu^i]$.

For contravariant vectors:

Let $\tilde{G} = [g_{ij}]$, where i is the row indicator and j is the column indicator. This means that the index of $\tilde{\lambda}$ has to be a column indicator, so we transpose $\tilde{\lambda}$. The index of μ is a row indicator, which matches with j -the column indicator of \tilde{G} , so we are good to proceed. Again, note that n is not necessarily 3.

$$\vec{\lambda} \cdot \vec{\mu} = \lambda^i g_{ij} \mu^j = (\lambda_1 \quad \dots \quad \lambda_n) \begin{pmatrix} g_{11} & \dots & g_{1n} \\ \vdots & \ddots & \vdots \\ g_{n1} & \dots & g_{nn} \end{pmatrix} \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix} = \tilde{\lambda}^\top \tilde{G} \tilde{\mu}$$

For covariant vectors:

We can find express the inner product of covariant vectors in matrix multiplication, using the form we have found for contravariant products. Consider two covariant vectors $\tilde{\lambda}^* = [\lambda_i]$ and $\tilde{\mu}^* = [\mu_i]$ and the inverse metric tensor $\tilde{G}^{-1} = [g_{ij}]^{-1} = [g^{ij}]$ (recall that $g_{ij}g^{jk} = \delta_i^k$). Similar to the contravariant vector case, but now with $\tilde{\lambda} \cdot \tilde{\mu} = g^{ij}\lambda_i\mu_j$:

$$\tilde{\lambda}^* \cdot \tilde{\mu}^* = (\lambda_1 \quad \dots \quad \lambda_n) = \begin{pmatrix} g^{11} & \dots & g^{1n} \\ \vdots & \ddots & \vdots \\ g^{n1} & \dots & g^{nn} \end{pmatrix} \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix} = \tilde{\lambda}^{*\top} \tilde{G}^{-1} \tilde{\mu}$$

As an little aside, to write $\tilde{\lambda}^*$ in terms of the metric tensor matrix and the contravariant vector, we simply remember that $\lambda_i = g_{ij}\lambda^j$. So,

$$\begin{aligned} \tilde{\lambda}^* &= \tilde{G} \tilde{\lambda} \\ \tilde{\lambda} &= \tilde{G}^{-1} \tilde{\lambda}^* \end{aligned}$$

In matrix form, the above two equalities can be written as:

$$\begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix} = \begin{pmatrix} g_{11} & \dots & g_{1n} \\ \vdots & \ddots & \vdots \\ g_{n1} & \dots & g_{nn} \end{pmatrix} \begin{pmatrix} \lambda^1 \\ \vdots \\ \lambda_n \end{pmatrix}$$

and

$$\begin{pmatrix} \lambda^1 \\ \vdots \\ \lambda^n \end{pmatrix} = \begin{pmatrix} g^{11} & \dots & g^{1n} \\ \vdots & \ddots & \vdots \\ g^{n1} & \dots & g^{nn} \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}$$

respectively.

Metric tensor and inverse metric tensor

As mentioned, $[g^{ij}] = [g_{ij}]^{-1}$. In matrix form:

$$\tilde{G}^{-1} \tilde{G} = \begin{pmatrix} g^{11} & \dots & g^{1n} \\ \vdots & \ddots & \vdots \\ g^{n1} & \dots & g^{nn} \end{pmatrix} \begin{pmatrix} g_{11} & \dots & g_{1n} \\ \vdots & \ddots & \vdots \\ g_{n1} & \dots & g_{nn} \end{pmatrix} = \text{diag}(1, \dots, 1) = \tilde{I}$$

Example 3.8. Find the inverse metric tensor g^{ij} in spherical coordinates, given the metric tensor g_{ij} is:

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

By definition, $[g^{ij}] = [g_{ij}]^{-1}$. So taking the inverse of $[g_{ij}]$ will give us $[g^{ij}]$. But notice that $[g_{ij}]$ is diagonal, so g^{ij} is simply the reciprocal of g_{ij} , if $g_{ij} \neq 0$, and 0 otherwise. This gives

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

3.7 Coordinate transformation

We have already seen the usage of multiple (more or less conventional) coordinate systems in the three-dimensional space. In general n -dimensional spaces, the options for coordinate systems are almost unlimited. It is crucial for us, as relativists, to develop for rigorous procedure to transform vector and tensor components as we move from one coordinate system to another. Note that vector and tensor **components** change under general coordinate transformation. It is common in physics to associate a vector or a tensor with a physical quantity, and coordinate system, or reference frame, etc., with perspectives. It makes sense that what we see might differ at different perspectives, but the physical quantity itself is independent of how we view it. In other words, vectors and tensors do not change under general coordinate transformations, only their components do.

In this subsection, we will learn how to transform between arbitrary coordinate systems. We will also learn how vector and tensor components transform, as well as what vectors and tensors are.

3.7.1 Vectors

We as physicists often think of vectors as an object with magnitude and direction. But this definition is incomplete, as it disregards a reference frame in which the vector is viewed. As we have mentioned in the introduction to this section, under general coordinate transformations, vectors do not change, but their components do, since the basis set of the coordinate system changes. For instance,

insert figure 10 here!!!

insert figure 11 here!!!

In the above two figures, the vector $\vec{\lambda}$ is the same. However, one frame is rotated by θ from another. In frames (S) and (S') , $\vec{\lambda}$ will have different components, but the “physical object” $\vec{\lambda}$ does not change. For instance, we can calculate the norm of $\vec{\lambda}$, $||\vec{\lambda}||$, and find that it is the same in both frames. In fact, **scalars are invariant under general coordinate transformations**.

Next, let’s talk **notations**. Let’s assume, without loss of generality, that we are only dealing with contravariant vectors here (the rules we will establish

shortly will apply to covariant vectors as well). Consider $\vec{\lambda}$ in frame (S) . In index-notation, we can write $\vec{\lambda} = \{\lambda^i\}$. Consider a different frame (a “primed” frame), denoted (S') . We will denote the components of $\vec{\lambda}$ in (S') as $\lambda^{i'}$. Just to get us familiar with the new notation, let’s go through an example:

Example 3.9. Suppose that $\vec{\lambda}$ is described two coordinate systems: spherical and cylindrical. Let the “unprimed” coordinate system be the spherical coordinate system. Then,

$$\begin{aligned}\{u^i\} &= (r, \theta, \phi) \\ \{u^{i'}\} &= (\rho, \phi, z).\end{aligned}$$

As simple as that!

Back to (S) and (S') . Since these are two different coordinate systems, the basis sets and metric tensors are also different:

$$\begin{aligned}(S) : \vec{e}_i &= \frac{\partial \vec{r}}{\partial u^i}, \quad \vec{e}^i = \vec{\nabla} u^i, \quad g_{ij} = \vec{e}_i \cdot \vec{e}_j \\ (S') : \vec{e}_{i'} &= \frac{\partial \vec{r}}{\partial u^{i'}}, \quad \vec{e}^{i'} = \vec{\nabla} u^{i'}, \quad g_{i'j'} = \vec{e}_{i'} \cdot \vec{e}_{j'}.\end{aligned}$$

Now, since $\vec{\lambda}$ is the same in both (S) and (S') , it must be true that”

$$\vec{\lambda} = \lambda^i \vec{e}_i = \lambda^{i'} \vec{e}_{i'},$$

i.e., λ^i and \vec{e}_i must transform in a way that leaves $\vec{\lambda}$ unchanged. Consider $\vec{r} = \vec{r}(u^{i'}) = \vec{r}(u^i(u^j))$. By the chain rule:

$$\vec{e}_j = \frac{\partial \vec{r}}{\partial u^j} = \frac{\partial \vec{r}}{\partial u^{i'}} \frac{\partial u^{i'}}{\partial u^j} = \vec{e}^{i'} \frac{\partial u^{i'}}{\partial u^j}.$$

Let us define, for simplicity sake,

$$U_j^{i'} = \frac{\partial u^{i'}}{\partial u^j}$$

Note that there is nothing new here. $U_j^{i'}$ simply generalizes 9 partial derivatives (because we are in 3-dimensional space). Together as a matrix, $[U_j^{i'}]$ is nothing but the Jacobian matrix “representing” the transformation $(S) \rightarrow (S')$:

$$[U_j^{i'}] = \begin{pmatrix} \frac{\partial u^{1'}}{\partial u^1} & \frac{\partial u^{1'}}{\partial u^2} & \frac{\partial u^{1'}}{\partial u^3} \\ \frac{\partial u^{2'}}{\partial u^1} & \frac{\partial u^{2'}}{\partial u^2} & \frac{\partial u^{2'}}{\partial u^3} \\ \frac{\partial u^{3'}}{\partial u^1} & \frac{\partial u^{3'}}{\partial u^2} & \frac{\partial u^{3'}}{\partial u^3} \end{pmatrix}$$

So, the chain rule for \vec{e}_j becomes the general coordinate transformation rule for the natural basis sets $\vec{e}_{i'} \rightarrow \vec{e}_i$.

$$\vec{e}_j = U_j^{i'} \vec{e}_{i'}$$

Next, since $\vec{\lambda} = \lambda^{i'} \vec{e}_{i'} = \lambda^j \vec{e}_j = \lambda^j U_j^{i'} \vec{e}_{i'}$, we obtain the general coordinate transformation rule for contravariant vector components $\lambda^i \rightarrow \lambda^{i'}$:

$$\boxed{\lambda^{i'} = U_j^{i'} \lambda^j}$$

We can, of course, conjecture that the transformation rule for contravariant vector components $\lambda^{i'} \rightarrow \lambda^i$ is:

$$\boxed{\lambda^i = U_{i'}^j \lambda^{i'}}$$

and we would be correct. The “proof” follows here. First, we look at the Jacobian matrix representing the transformation $(S') \rightarrow (S)$:

$$[U_{i'}^j] = \left[\frac{\partial u^j}{\partial u^{i'}} \right],$$

which is simply the inverse of $[U_j^{i'}]$. So, it follows that:

$$[U_k^{j'}][U_{i'}^k] = \text{diag}(1, \dots, 1) = [\delta_{i'}^{j'}] = [\delta_i^j] = \tilde{I},$$

and

$$[U_{i'}^k][U_j^{i'}] = \text{diag}(1, \dots, 1) = [\delta_j^k] = \tilde{I},$$

because matrices that are inverses of each other are commutative. Element-wise:

$$\boxed{U_{i'}^k U_j^{i'} = \delta_j^k = \delta_{j'}^{k'} = U_i^{k'} U_{j'}^i}$$

So we can now verify our little “conjecture:”

$$\begin{aligned} \lambda^{i'} &= U_j^{i'} \lambda^j \\ U_{i'}^k \lambda^{i'} &= U_j^{i'} U_{i'}^k \lambda^j \\ U_{i'}^k \lambda^{i'} &= \delta_j^k \lambda^j \\ U_{i'}^k \lambda^{i'} &= \lambda^k. \end{aligned}$$

We notice that nothing makes $U_j^{i'}$ more special than $U_{j'}^i$. Also, while we use contravariant vectors more often than covariant vectors, nothing makes contravariant vectors more special than covariant vectors. So if transformation rules apply to contravariant vectors, we should also expect them to work for covariant vectors:

$$\boxed{\begin{aligned} \lambda_{i'} &= U_{i'}^j \lambda_j \\ \lambda_i &= U_i^{j'} \lambda_{j'} \end{aligned}}$$

At this point it looks like we are inventing new mathematics by playing a little game of swapping indices, and it is reasonable for us to grow a little bit suspicious

of what we are doing here. But we can show that our “index-swapping intuition” is correct. Let us consider (again) a vector $\vec{\lambda} = \lambda_j \vec{e}^j = \{\lambda_j\}$, where \vec{e}^j is a vector in the dual basis:

$$\vec{e}^j = \vec{\nabla} u^j = \frac{\partial u^j}{\partial x} \hat{i} + \frac{\partial u^j}{\partial y} \hat{j} + \frac{\partial u^j}{\partial z} \hat{k}.$$

Let us apply to chain rule:

$$\frac{\partial u^j}{\partial x} = \frac{\partial u^j}{\partial u^{i'}} \frac{\partial u^{i'}}{\partial x}.$$

Therefore,

$$\begin{aligned} \vec{e}^j &= \vec{\nabla} u^j \\ &= \frac{\partial u^j}{\partial x} \hat{i} + \frac{\partial u^j}{\partial y} \hat{j} + \frac{\partial u^j}{\partial z} \hat{k} \\ &= \frac{\partial u^j}{\partial u^{i'}} \frac{\partial u^{i'}}{\partial x} \hat{i} + \frac{\partial u^j}{\partial u^{i'}} \frac{\partial u^{i'}}{\partial y} \hat{j} + \frac{\partial u^j}{\partial u^{i'}} \frac{\partial u^{i'}}{\partial z} \hat{k} \\ &= \sum_{i=1}^3 \frac{\partial u^j}{\partial u^{i'}} \frac{\partial u^{i'}}{\partial x} \hat{i} + \frac{\partial u^j}{\partial u^{i'}} \frac{\partial u^{i'}}{\partial y} \hat{j} + \frac{\partial u^j}{\partial u^{i'}} \frac{\partial u^{i'}}{\partial z} \hat{k} \\ &= \sum_{i=1}^3 \frac{\partial u^j}{\partial u^{i'}} \left(\frac{\partial u^{i'}}{\partial x} \hat{i} + \frac{\partial u^{i'}}{\partial y} \hat{j} + \frac{\partial u^{i'}}{\partial z} \hat{k} \right) \\ &= \sum_{i=1}^3 \frac{\partial u^j}{\partial u^{i'}} \vec{\nabla} u^{i'} \\ &= \sum_{i=1}^3 U_{i'}^j \vec{e}^{i'}. \end{aligned}$$

Or, in Einstein’s summation notation, the transformation rule for the dual basis vectors is:

$$\boxed{\vec{e}^j = U_{i'}^j \vec{e}^{i'}}$$

And the inverse transformation follows:

$$\boxed{\vec{e}^{j'} = U_{i'}^{j'} \vec{e}^{i'}}$$

Note that, the above relation is in fact true for any n -dimensional space. Our “proof” here is restricted to only 3-dimensional space. However, it is not so difficult to fix this slight issue. If we redefine the $\vec{\nabla}$ operator for n -dimensional space, then we can easily see that the new proof is identical to the proof above, except for the maximal value of n .

We easily obtain the transformation rules for the covariant vector components using the results above:

$$\vec{\lambda} = \lambda_{i'} \vec{e}^{i'} = \lambda_j \vec{e}^j = \lambda_j U_{i'}^j \vec{e}^{i'} = \lambda_{i'} U_j^{i'} \vec{e}^j.$$

Therefore,

$$\boxed{\begin{aligned} \lambda_{i'} &= U_{i'}^j \lambda_j \\ \lambda_j &= U_j^{i'} \lambda_{i'} \end{aligned}}$$

Below is a summary of what we have learned so far about general coordinate transformations:

Natural basis:	$\vec{e}_j = U_j^{i'} \vec{e}_{i'}$	$\vec{e}_{j'} = U_{j'}^i \vec{e}_i$
Dual basis:	$\vec{e}^j = U_{i'}^j \vec{e}^{i'}$	$\vec{e}^{j'} = U_i^{j'} \vec{e}^i$
Contravariants:	$\vec{\lambda}^{j'} = U_{i'}^{j'} \vec{\lambda}^i$	$\vec{\lambda}^j = U_{i'}^j \vec{\lambda}^{i'}$
Covariants:	$\vec{\lambda}_j = U_j^{i'} \vec{\lambda}_{i'}$	$\vec{\lambda}_{j'} = U_{j'}^i \vec{\lambda}_i$

Since components of a vector must transform according to the rules above under general coordinate transformations, we can turn these rules into a more general definition of a vector.

Definition 3.1. A vector is a quantity whose components transform contravariantly as $\lambda^{i'} = U_j^{i'} \lambda^j$ under a general coordinate transformation $u^{i'} = u^{i'}(u^j)$.

Remark 3.1. We are often interested in vector fields, a collection of vectors at different points in space whose components depend on the coordinates: $\lambda^i = \lambda^i(u^j)$. Since vectors have to transform correctly, we require that at any point P , we require that $\lambda^{i'} = U_j^{i'} \lambda^j$ be true for a vector field (defined on λ^i) to be defined.

Remark 3.2. Not all 3-tuple of functions are vectors. Consider a 3-tuple of coordinates. Let $\lambda^i = u^i$ and $\lambda^{j'} = u^{j'}$, where $u^{j'} = u^{j'}(u^i)$. For λ^i to be make a vector field, it is required that $\lambda^{i'} = U_j^{i'} \lambda^j$. This implies that $u^{i'} = U_j^{i'} u^j$, with $U_j^{i'} = \partial u^{i'} / \partial u^j$. However, we immediately see that this is **not true in general**:

$$\boxed{u^{i'} \neq \frac{\partial u^{i'}}{\partial u^j} u^j}$$

But rather, $u^{i'} = u^{i'}(u^j) = f(u^j)$. So, coordinates themselves do not make vectors in general, because they, as components, do not transform correctly under general coordinate transformations. This is why we never use g_{ij} or g^{ij} to lower or raise the index of u^i or u_i , i.e., $u^i \neq g^{ij} u_j$, and $u_j \neq g_{ij} u^i$, in general.

There are exemptions to the “general rule,” of course. For instance, consider **linear transformations**, where $u^{j'} = u^{j'}(u^i) = C_i^{j'} u^i$, where $C_i^{j'}$ are

constants. What this means is that the new coordinates are simply linear combinations of the old coordinates. It follows that

$$\frac{\partial u^{j'}}{\partial u^k} = C_i^{j'} \frac{\partial u^i}{\partial u^k} = C_i^{j'} \delta_k^i = C_k^{j'}.$$

If we let $k = i$ (we're just swapping dummy indices here), then $C_i^{j'} = \partial u^{j'} / \partial u^i = U_i^{j'}$, i.e., $u^{j'} = U_i^{j'} u^i$ under linear transformations. Therefore, u^i transform correctly, and in this case the coordinates u^i can form a vector, $\{u^i\}$.

Remark 3.3. Properly speaking we can define vectors with respect to a particular class of transformations. It is possible for something to be a vector with respect to a class of transformation, but not a vector with respect to another. Such a class of transformation (default) is the general coordinate transformations.

Example 3.10. Can differentials of coordinates (du^i) form a vector?

The solution is actually quite simple. Without loss of generality, let us consider a hypothetical vector $\{du^i\}$ in 3-dimensional space. $\{du^1, du^2, du^3\}$. Applying the chain rule:

$$du^{i'} = \frac{\partial u^{i'}}{\partial u^j} du^j = U_j^{i'} du^j.$$

We see that du^i transform correctly under general coordinate transformations. Therefore, du^i can make a vector.

Example 3.11. Find $[U_i^{j'}]$ for a coordinate transformation from Cartesian to spherical in flat 3-dimensional space. Let $\{u^j\} = \{x, y, z\}$ and $\{u^{j'}\} = \{r, \theta, \phi\}$. We find the Jacobian matrix:

$$[U_i^{j'}] = \begin{pmatrix} \frac{\partial r}{\partial x} & \frac{\partial r}{\partial y} & \frac{\partial r}{\partial z} \\ \frac{\partial \theta}{\partial x} & \frac{\partial \theta}{\partial y} & \frac{\partial \theta}{\partial z} \\ \frac{\partial \phi}{\partial x} & \frac{\partial \phi}{\partial y} & \frac{\partial \phi}{\partial z} \end{pmatrix} = \frac{1}{r} \begin{pmatrix} r \sin \theta \cos \phi & r \sin \theta \sin \phi & r \cos \theta \\ \cos \theta \cos \phi & \cos \theta \sin \phi & -\sin \theta \\ -\frac{\sin \phi}{\sin \theta} & \frac{\cos \phi}{\sin \theta} & 0 \end{pmatrix}$$

Example 3.12. Consider $\vec{\lambda} = (1, 0, 0)^\top = \{\lambda^{i'}\}$ in Cartesian coordinates. What are the components of $\vec{\lambda}$ in spherical coordinates?

With our developed procedure, we should be able to obtain the answer with little thinking:

$$\vec{\lambda} = \{\lambda^{i'}\} = \begin{pmatrix} \lambda^{1'} \\ \lambda^{2'} \\ \lambda^{3'} \end{pmatrix} = [U_j^{i'} \lambda^j] = \frac{1}{r} \begin{pmatrix} r \sin \theta \cos \phi & r \sin \theta \sin \phi & r \cos \theta \\ \cos \theta \cos \phi & \cos \theta \sin \phi & -\sin \theta \\ -\frac{\sin \phi}{\sin \theta} & \frac{\cos \phi}{\sin \theta} & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix},$$

which gives

$$\begin{pmatrix} \lambda^{1'} \\ \lambda^{2'} \\ \lambda^{3'} \end{pmatrix} = \frac{1}{r} \begin{pmatrix} r \sin \theta \cos \phi \\ \cos \theta \cos \phi \\ -\frac{\sin \phi}{\sin \theta} \end{pmatrix}.$$

In the natural basis of spherical coordinates,

$$\vec{\lambda} = \sin \theta \cos \phi \vec{e}_r + \frac{1}{r} \cos \theta \cos \phi \vec{e}_\theta - \frac{\sin \phi}{r \sin \theta} \vec{e}_\phi.$$

We have said before that the “physical quantity” $\vec{\lambda}$ does not change. We can test this by finding $||\vec{\lambda}||$ in Cartesian and spherical coordinates and compare the results. In Cartesian, the metric tensor is $g_{ij} = \delta_i^j$, so

$$||\vec{\lambda}||_{\text{Cartesian}} = \sqrt{g_{ij} \lambda^i \lambda^j} = \sqrt{1^2 + 0^2 + 0^2} = 1,$$

as expected. The metric tensor $g_{i'j'}$ of spherical coordinate, in matrix form, is:

$$[g_{i'j'}] = \begin{pmatrix} 1 & 0 & 0 \\ 0 & r^2 & 0 \\ 0 & 0 & r^2 \sin^2 \theta \end{pmatrix}.$$

It follows that

$$\begin{aligned} ||\vec{\lambda}||_{\text{Spherical}} &= \sqrt{g_{i'j'} \lambda^{i'} \lambda^{j'}} = g_{1'1'} (\lambda^{1'})^2 + g_{2'2'} (\lambda^{2'})^2 + g_{3'3'} (\lambda^{3'})^2 \\ &= \sqrt{(\sin \theta \cos \phi)^2 + r^2 \left(\frac{1}{r} \cos \theta \cos \phi \right)^2 + r^2 \sin^2 \theta \left(-\frac{\sin \phi}{r \sin \theta} \right)^2} \\ &= \sqrt{(\sin \theta \cos \phi)^2 + (\cos \theta \cos \phi)^2 + \sin^2 \phi} \\ &= \sqrt{\cos^2 \phi + \sin^2 \phi} \\ &= 1. \end{aligned}$$

Not surprisingly, $||\vec{\lambda}_{\text{Cartesian}}|| = ||\vec{\lambda}_{\text{Spherical}}|| = 1$.

Example 3.13. Find $U_i^{j'}$ for a rotation by an angle ϕ around the z -axis in Cartesian coordinates. Suppose $\vec{a} = \{a^i\} = (1, 1, 0)^\top$. What is $\vec{a} = \{a^{i'}\}$ after a rotation by ϕ ?

insert figure 12 here!!!

From linear algebra, we know that a rotation by ϕ around the z -axis is given by:

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} \cos \phi & -\sin \phi & 0 \\ \sin \phi & \cos \phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

So,

$$[U_i^{j'}] = \begin{pmatrix} \cos \phi & -\sin \phi & 0 \\ \sin \phi & \cos \phi & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Given that $\vec{a} = \{a^i\} = (1, 1, 0)^\top$,

$$\begin{pmatrix} \lambda^{1'} \\ \lambda^{2'} \\ \lambda^{3'} \end{pmatrix} = \begin{pmatrix} \cos \phi & -\sin \phi & 0 \\ \sin \phi & \cos \phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} \cos \phi - \sin \phi \\ \sin \phi + \cos \phi \\ 0 \end{pmatrix}$$

Notice that because the coefficients of $[U_i^{j'}]$ are constants, the coordinates $\{x, y, z\}$ can make a vector. But note (again) that this is not true in general.

Next, we can find $||\vec{\lambda}||$ in both coordinate systems again and show that they are equal. In Cartesian coordinates, $\{\lambda^i\} = (1, 1, 0)^\top$. So, $||\vec{\lambda}|| = \sqrt{2}$. In the rotated frame:

$$||\vec{\lambda}|| = \sqrt{g_{i'j'} \lambda^{i'} \lambda^{j'}}.$$

We seem to be stuck here, since we have yet found the form of the new metric $g_{i'j'}$. However, we observe that $g_{i'j'} = \vec{e}_{i'} \cdot \vec{e}_{j'}$, which is just either 1 (if $i = j$) or 0 (if $j \neq i$), because the new coordinate system is just the “tilted” version of the old one. So we expect that new metric tensor is the same as the old metric tensor, which is nothing but $g_{i'j'} = \text{diag}(1, 1, 1)$. We can proceed to calculate the norm:

$$\begin{aligned} ||\vec{\lambda}|| &= \sqrt{(\cos \phi - \sin \phi)^2 + (\sin \phi + \cos \phi)^2 + 0^2} \\ &= \sqrt{2(\sin^2 \phi + \cos^2 \phi)} \\ &= \sqrt{2}, \end{aligned}$$

which is of course what we found earlier.

The above example motivates what we will discuss next. We have seen how vector components transform under general coordinate transformations, and how their norms (scalar) remain invariant. However, we start to question how the metric tensor transforms, in general. In the above example, the only reason we was able to calculate the norm of $\vec{\lambda}$ was that we could guess the form of the metric tensor, using some cleverness. In general, though, we will not be able to solve this problem without knowing exactly the form of the metric. Our next task, as apparent as it is now, is to find out how metric tensors in particular and tensors in general transform under general coordinate transformations.

3.7.2 Tensors

We should first ask ourselves what tensors are. While a rigorous definition of a tensor will come shortly in this section (hint: it will resemble that of a vector, i.e., based on how their components transform under general coordinate transformations), we can first think of tensors as generalizations of vectors, i.e., with more than one index and hence multi-dimensional. Consider a physical object like a balloon which stretches and compresses multi-dimensionally when some force is applied to it. We call this “response” the stress tensor, which has 9 components in 3-dimensional space: $F_{xx}, F_{xy}, F_{xz}, F_{yx}, F_{yy}, F_{yz}, F_{zx}, F_{zy}, F_{zz}$.

Definition 3.2. A tensor is a multi-component quantity whose components transform as contravariant or covariant vector components.

Example 3.14. Find the condition for $\tau^{ij}{}_k$ to be a tensor.

Well, a tensor has to transform correctly, so we can play a little game of index-replacing:

$$\tau^{i'j'}{}_{k'} = U_m^{i'} U_n^{j'} U_p^k \tau^{mn}{}_p$$

under a general coordinate transformation $u^{i'} \rightarrow u^{i'}(u^j)$.

Example 3.15. Show that g_{ij} is a tensor.

g_{ij} is a tensor, in fact, by definition:

$$g_{i'j'} = U_{i'}^k \vec{e}_k \cdot U_{j'}^l \vec{e}_l = U_{i'}^k U_{j'}^l \vec{e}_k \cdot \vec{e}_l = U_{i'}^k U_{j'}^l g_{kl}$$

Similarly, the inverse g^{ij} is also a tensor:

$$g^{k'l'} = U_i^{k'} \vec{e}^i \cdot U_j^{l'} \vec{e}^j = U_i^{k'} U_j^{l'} \vec{e}^i \cdot \vec{e}^j = U_i^{k'} U_j^{l'} g^{ij}$$

Definition 3.3. A tensor $\tau^{abc\dots}_{def\dots}$ is of type (r, s) when it has r contravariant components and s covariant components.

Example 3.16. g_{ij} is a type $(0, 2)$ tensor. g^{ij} is a type $(2, 0)$ tensor.

Vectors are just a special type of tensors. Contravariant vectors λ^i are type $(1, 0)$ tensors, while covariant vectors λ_i are type $(0, 1)$ tensors.

Remark 3.4. The transformation $U_j^{i'}$ is NOT a tensor. Indeed, it does not make sense to talk about how a transformation transforms.

Example 3.17. Write $g_{i'j'} = U_{i'}^k U_{j'}^l g_{kl}$ as a matrix equation.

We let $\tilde{G} = [g_{ij}]$, $\tilde{G}' = [g_{i'j'}]$, and $\hat{U} = \tilde{U}^{-1} = [\partial u^k / \partial u^{i'}]$. To create a matrix equation, we have to be aware of when an index indicate a row or column. The correct “row-column-sensitive” order expression is:

$$[g_{i'j'}] = [U_{i'}^k][g_{kl}][U_{j'}^l],$$

or equivalently,

$$\boxed{\tilde{G}' = \hat{U}^\top \tilde{G} \hat{U}}$$

Example 3.18. Using the transformation rules for tensors, obtain $g_{i'j'}$ for a rotation by ϕ around the z -axis in 3-dimensional space.

Recall that

$$[U_i^{j'}] = \begin{pmatrix} \cos \phi & -\sin \phi & 0 \\ \sin \phi & \cos \phi & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

and that $g_{ij} = \text{diag}(1, 1, 1)$ in 3-dimensional Cartesian coordinate system. To find \tilde{G}' , we first have to find $U_{i'}^k$, the inverse transformation. We expect that $U_{i'}^k$ is simply a rotating by $-\phi$ (since this intuitively “reverses” what $U_i^{k'}$ does). But we, as always, find $U_{i'}^k$ mechanically:

$$\hat{U} = \left[\frac{\partial u^k}{\partial u^{i'}} \right] = \tilde{U}^{-1} = \begin{pmatrix} \cos \phi & -\sin \phi & 0 \\ \sin \phi & \cos \phi & 0 \\ 0 & 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} \cos \phi & \sin \phi & 0 \\ -\sin \phi & \cos \phi & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

as anticipated. Using the result of the previous example,

$$\begin{aligned} \tilde{G}' &= \hat{U}^\top \tilde{G} \hat{U} \\ &= \begin{pmatrix} \cos \phi & \sin \phi & 0 \\ -\sin \phi & \cos \phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \cos \phi & -\sin \phi & 0 \\ \sin \phi & \cos \phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \end{aligned}$$

just as guessed. So our intuition was correct after all. But we realize an important fact: the metric tensor can remain the same after a non-trivial transformation (in this case the transformation is a rotating around the z -axis). We also notice that the metric tensors are the same under a transformation if the transformation matrix is **orthogonal**, i.e., its inverse the equal to its transpose:

$$\hat{U} = \tilde{U}^{-1} = \tilde{U}^\top.$$

Remark 3.5. Only tensors of type (r, s) with $r + s \leq 2$ can be displayed as matrices and/or matrix multiplications. It is not possible to write τ_{kl}^{ij} as a matrix.

3.8 Scalars

As we have seen from time to time, scalars are invariant under general coordinate transformations. As tensors, scalars are of type $(0,0)$, i.e., they have open indices.

Example 3.19. Show that the magnitude of a vector is a scalar.

Consider $\vec{a} = a^i = a^{i'}$ in two different coordinate systems. The magnitude of $\vec{a} = ||\vec{a}||$ is a scalar if:

$$\vec{a} \cdot \vec{a} = a^i a_i = a^{i'} a_{i'}.$$

This can be readily shown, since we already know the transformation rules for vector components:

$$\begin{aligned} a^{i'} a_{i'} &= (U_j^{i'} a^j) (U_{i'}^k a_k) \\ &= U_j^{i'} a^j U_{i'}^k a_k \\ &= \delta_j^k a_k a^j \\ &= a^k a_k \\ &= a^i a_i. \end{aligned}$$

Therefore, $||\vec{a}||$ is a scalar.

Example 3.20. Show that the line element $ds^2 = g_{ij} du^i du^j$ is a scalar.

Similar to the previous example, we have to show that:

$$g_{i'j'} du^{i'} du^{j'} = g_{ij} du^i du^j.$$

We simply apply the transformation rule to each factor and play a little game of index-arranging:

$$\begin{aligned} g_{i'j'} du^{i'} du^{j'} &= (U_{i'}^k U_{j'}^l g_{kl}) (U_m^{i'} du^m) (U_n^{j'} du^n) \\ &= U_{i'}^k U_{j'}^l U_m^{i'} U_n^{j'} g_{kl} du^m du^n \\ &= \delta_m^k \delta_n^l g_{kl} du^m du^n \\ &= g_{kl} du^k du^l \\ &= g_{ij} du^i du^j. \end{aligned}$$

Therefore, $ds^2 = g_{ij} du^i du^j$ is a scalar.

4 Flat spacetime

Recall that in space, we require 3 coordinates. It makes sense that in order to describe spacetime, we need an extra coordinate to describe time, in which case, we will switch from Roman-character indices ($i, j, k, \dots = 1, 2, 3$) to Greek-character indices ($\mu, \nu, \sigma, \dots = 0, 1, 2, 3$). We also tend to (by an unwritten convention and of course not mandatory) write vectors as capital letters to distinguish 4-vectors from 3-vectors. For example, below are a number of ways we express a 4-vector:

$$X^\mu = \{X^0, X^1, X^2, X^3\} = (X^0, \vec{X}) = (X^0, X^i).$$

Notice that the “hat” notation is reserved for 3-vectors.

4.1 Special Relativity

Coordinate transformations in special relativity are Lorentz transformations under which the spacetime interval (in Cartesian coordinates)

$$ds^2 = c^2 dt^2 + dx^2 + dy^2 + dz^2 = (dx^0)^2 + (dx^1)^2 + (dx^2)^2 + (dx^3)^2,$$

which gives “physical distances” in spacetime, is invariant. We also recall that

$$ds^2 = g_{ij} du^i du^j,$$

we can “read off” the metric from the spacetime interval as $\text{diag} = (1, -1, -1, -1)$. This is referred to as the Minkowski metric (flat 4-dimensional spacetime in Cartesian coordinates), denoted as $\eta_{\mu\nu}$:

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

Because any frame in special relativity can be connected to the original frame by a Lorentz transformation, i.e.,

$$ds^2 = ds'^2 = (dx^{0'})^2 + (dx^{1'})^2 + (dx^{2'})^2 + (dx^{3'})^2,$$

The Minkowski metric (in Cartesian coordinates) is the same for any reference frame:

$$[\eta_{\mu'\nu'}] = [\eta_{\mu\nu}] = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}.$$

This says that:

$$ds^2 = \eta_{\mu\nu} dX^\mu dX^\nu = \eta_{\mu'\nu'} dX^{\mu'} dX^{\nu'}$$

For general coordinate systems in three dimensions, the metric $g_{ij} \neq \text{diag}(1, 1, 1)$. In these cases, the Minkowski metric has the following form:

$$[\eta_{\mu\nu}] = \begin{pmatrix} 1 & 0 \\ 0 & -[g_{ij}] \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & & & \\ 0 & & -[g_{ij}] & \\ 0 & & & \end{pmatrix}.$$

More generally, however, spacetime can be curved, and the metric tensor of the three-dimensional space can be non-Cartesian. In these cases, we simply replace $\eta_{\mu\nu}$ by $g_{\mu\nu}$. It follows that:

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

in general. The “moral” here is that $\eta_{\mu\nu}$ is reserved for flat Minkowski spacetime.

Same as before, the metric tensor encodes all “physical properties” of the Minkowski spacetime and allows us to transform contravariant components into covariant components and vice versa. For instance, consider a contravariant vector $\lambda^\mu = (\lambda^0, \lambda^1, \lambda^2, \lambda^3) = (\lambda^0, \vec{\lambda})$. The covariant vector is given by simply “lowering” the index of the contravariant vector using the Minkowski metric tensor:

$$\lambda_\mu = \eta_{\mu\nu} \lambda^\nu = (\lambda_0, \lambda_1, \lambda_2, \lambda_3).$$

The inverse transformation has the same form:

$$\lambda^\mu = \eta^{\mu\nu} \lambda_\nu.$$

Note that if we use the Minkowski metric tensor in Cartesian coordinates, then

$$\lambda_\mu = \eta_{\mu\nu} \lambda^\nu = (\lambda_0, \lambda_1, \lambda_2, \lambda_3) = (\lambda^0, -\lambda^1, -\lambda^2, -\lambda^3).$$

We observe that

$$\begin{aligned} \lambda^0 &= \lambda_0 \\ \lambda^i &= -\lambda_i. \end{aligned}$$

We also observe that the inverse Minkowski metric tensor is Minkowski metric tensor itself:

$$[\eta^{\mu\nu}] = [\eta_{\mu\nu}]^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

Inner products of vectors follow the same rule as before:

$$\mathbf{a} \cdot \mathbf{b} = a^\mu b_\mu = a_\mu b^\mu = \eta_{\mu\nu} a^\mu b^\nu = \eta^{\mu\nu} a_\mu b_\nu$$

On a more practical side of things, we have to be careful with the signs that go with the terms when we solve problems:

$$\begin{aligned}\mathbf{a} \cdot \mathbf{b} &= a^1 b_1 + a^2 b_2 + a^2 b_2 + a^3 b_3 \\ &= a_1 b_1 - a_2 b_2 - a_2 b_2 - a_3 b_3 \\ &= a^1 b^1 - a^2 b^2 - a^2 b^2 - a^3 b^3.\end{aligned}$$

At this point, we might wonder what the basis vectors \vec{e}_μ in the flat, 4-dimensional Minkowski spacetime in Cartesian coordinates are. We can, of course, find them, from the metric tensor $\eta_{\mu\nu}$. However, one thing for sure: they are not going to be the basis vectors $\hat{i}, \hat{j}, \hat{k}$. In fact, the basis vectors that make up $\eta_{\mu\nu}$ will likely contain imaginary parts, since $\vec{e}_i \cdot \vec{e}_j = -1$ for $i = j$. The point is: we don't really need the basis vectors once we already have the metric, because after all, almost all of our computations involve the metric tensor.

4.1.1 The Lorentz Transformations, revisited

Recall that Lorentz transformations are (not necessarily strictly linear) transformations from one **inertial frame** (S) to another **inertial frame** (S').

The most general Lorentz transformations, also referred to collectively as “Poincaré transformations” include:

- Lorentz boost: inertial frames with relative velocity v .
- Translation: origins don't coincide at $t' = t = 0$.
- Spatial rotation
- Spatial inversion: or parity transformations ($x' = -x$)
- Time reversal: $t' = -t$.

We can also make further distinctions among the listed transformations:

- Inhomogeneous Lorentz transformations: have non-trivial translation.
- Homogeneous Lorentz transformations: no translations.
- Improper Lorentz transformations: with time reversal or parity.
- Proper Lorentz transformation: no parity or time reversal.

We can first look at **homogeneous, proper Lorentz transformations with no rotations**, i.e., Lorentz boosts. These are probably the simplest of Lorentz transformations that we encountered in introductory special relativity. A classic example is the Lorentz boost along the x -axis:

$$\begin{pmatrix} a^{0'} \\ a^{1'} \\ a^{2'} \\ a^{3'} \end{pmatrix} = \begin{pmatrix} \gamma & -\beta\gamma & 0 & 0 \\ -\beta\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a^0 \\ a^1 \\ a^2 \\ a^3 \end{pmatrix}$$

where the transformation matrix element $U_j^{i'}$ in 3-dimensional space is now generalized to 4-dimensional spacetime:

$$X_\nu^{\mu'} = \frac{\partial a^{\mu'}}{\partial a^\nu}$$

Specifically for Lorentz boosts, we let the general $X_\nu^{\mu'}$ be $\Lambda_\nu^{\mu'}$. This slight change in notation gives:

$$[\Lambda_\nu^{\mu'}] = \left[\frac{\partial a^{\mu'}}{\partial a^\nu} \right] = \begin{pmatrix} \gamma & -\beta\gamma & 0 & 0 \\ -\beta\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

We notice immediately that $\Lambda_\nu^{\mu'}$ is constant for all μ, ν . This means (i) Lorentz transformations are linear transformations, and (ii) Cartesian coordinates X^μ can form the components of a vector under Lorentz transformations because

$$X^{\mu'} = \Lambda_\nu^{\mu'} X^\nu$$

is obeyed. While this concept might seem foreign, we realize we have seen this many times before in special relativity, only in a slightly different form:

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

which gives us the very first Lorentz-Einstein equations that we are familiar with:

$$\begin{cases} ct' = \gamma(ct - \beta x) \\ x' = \gamma(x - \beta ct) \\ y' = y \\ z' = z. \end{cases}$$

Keep in mind that the equalities above are only true because $\Lambda_\nu^{\mu'} \Lambda_\sigma^\nu = \delta_\sigma^\mu$, and are not true in general, because coordinates don't form vectors in general.

We can find the inverse of $\Lambda_\nu^{\mu'}$ in two ways: the intuitive way, and the mechanical way. Intuitively, $\Lambda_\nu^{\mu'}$, or the inverse transformation, undoes what $\Lambda_\nu^{\mu'}$, so it has to be a boost in the opposite direction ($v \rightarrow -v$ therefore $\beta \rightarrow -\beta$). So if $\Lambda_\nu^{\mu'}$ is an x -boost, then its inverse is simply:

$$[\Lambda_\nu^{\mu}] = \begin{pmatrix} \gamma & \beta\gamma & 0 & 0 \\ \beta\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

which not surprisingly gives us the inverse Lorentz-Einstein transformation equations:

$$\begin{cases} ct = \gamma(ct' + \beta x) \\ x = \gamma(x + c\beta t) \\ y = y \\ z = z. \end{cases}$$

A reasonable next step is to look at **proper homogeneous Lorentz transformation with rotation(s)**. Recall from linear algebra that such a transformation is simply a composition of multiple, more elementary transformations (pure boost and/or pure rotation). For example, a pure rotation an angle ϕ about the z -axis has the following matrix:

$$[\Lambda_\nu^{\mu'}] = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \phi & \sin \phi & 0 \\ 0 & -\sin \phi & \cos \phi & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

To think how to do about how to construct a transformation matrix has includes both a boost and a rotation, we can first do a gentle example.

Example 4.1. Find the transformation matrix for a Lorentz boost in the y -axis in two ways: intuition and composition of rotation by $\pi/2$ and boost along the x -direction.

We can easily predict that the matrix for a boost in y should be almost the same as that for a boost x , except for the locations of the terms in the matrix. We know this because the x -direction is no more special than the y -direction. So,

$$[\Lambda_\nu^{\mu'}] = \begin{pmatrix} \gamma & 0 & -\beta\gamma & 0 \\ 0 & 0 & 0 & 0 \\ -\beta\gamma & 0 & \gamma & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

We should get the same matrix by doing the “composition method.” We can simply think of a boost in the y -direction as a rotating by $\pi/2$, followed by a boost along the x -direction, then followed by a rotation by $-\pi/2$. A final transformation matrix is a product of the list matrix, in right-to-left order:

$$\begin{aligned} [\Lambda_\nu^{\mu'}] &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \gamma & \beta\gamma & 0 & 0 \\ \beta\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} \gamma & 0 & -\beta\gamma & 0 \\ 0 & 0 & 0 & 0 \\ -\beta\gamma & 0 & \gamma & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \end{aligned}$$

as expected.

The moral of the above example is to show that any Lorentz boost along an arbitrary direction can be found as a combination of a boost along the x -direction and a spatial rotation:

$$[\Lambda_\nu^{\mu'}] = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \phi & -\sin \phi & 0 \\ 0 & \sin \phi & \cos \phi & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \gamma & \beta\gamma & 0 & 0 \\ \beta\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos \phi & \sin \phi & 0 \\ 0 & -\sin \phi & \cos \phi & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

4.1.2 A curiosity about the Lorentz boosts

We can actually make Lorentz boosts look like rotation using hyperbolic functions:

$$\begin{aligned} \sinh(\alpha) &= \frac{e^\alpha - e^{-\alpha}}{2} \\ \cosh(\alpha) &= \frac{e^\alpha + e^{-\alpha}}{2} \\ \operatorname{sech}(\alpha) &= \frac{1}{\cosh(\alpha)} \\ \operatorname{csch}(\alpha) &= \frac{1}{\sinh(\alpha)} \\ \tanh(\alpha) &= \frac{\sinh(\alpha)}{\cosh(\alpha)} \\ \coth(\alpha) &= \frac{1}{\tanh(\alpha)}, \end{aligned}$$

which have the following identities:

$$\begin{aligned} \cosh^2(\alpha) - \sinh^2(\alpha) &= 1 \\ 1 - \tanh^2(\alpha) &= \operatorname{sech}^2(\alpha). \end{aligned}$$

If we let $\beta = \tanh(\phi)$, where the angle ϕ is called “rapidity,” then it follows that:

$$\gamma = \frac{1}{\sqrt{1-\beta^2}} = \frac{1}{\sqrt{1-\tanh^2(\phi)}} = (\operatorname{sech}(\phi))^{-1} = \cosh(\phi),$$

and

$$\gamma\beta = \sinh(\phi).$$

So the Lorentz boost $[\Lambda_\nu^{\mu'}]$ becomes:

$$[\Lambda_\nu^{\mu'}] = \begin{pmatrix} \cosh(\phi) & -\sinh(\phi) & 0 & 0 \\ -\sinh(\phi) & \cosh(\phi) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

4.1.3 The Poincaré transformations

The general form of Poincaré transformations are given by:

$$a^{\mu'} = \Lambda_\nu^{\mu'} a^\nu + b^{\mu'}$$

where b^μ is a constant, i.e., $\frac{\partial b^{\mu'}}{\partial x^\nu} = 0$. We notice that, by the form of the equation, these are “affine” transformations, or equivalent a linear transformation (or linear map) plus a shift/translation. What we mean by “affine” is that, suppose we take $\frac{\partial a^{\mu'}}{\partial a^\nu}$:

$$\frac{\partial a^{\mu'}}{\partial a^\nu} = \Lambda_\nu^{\mu'},$$

which is constant, and “looks like a Lorentz transformation.” With the chain rule:

$$\Lambda_\nu^{\mu'} \Lambda_{\sigma'}^\nu = \frac{\partial a^{\mu'}}{\partial a^\nu} \frac{\partial a^\nu}{\partial a^{\sigma'}} = \delta_{\sigma'}^{\mu'},$$

which is what we have seen in Lorentz transformations.

A key feature of the Lorentz transformation is that it is a **linear map**, and that

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

where

$$[\eta_{\mu\nu}] = [\eta_{\mu'\nu'}] = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix},$$

i.e., the Lorentz transformations preserve the Minkowski metric in Cartesian coordinates. Now, if we take the differentials of the Poincaré transformation equation:

$$da^{\mu'} = \Lambda_{\sigma}^{\mu'} da^{\sigma},$$

and plug this into the ds^2 equation above, then

$$\begin{aligned} \eta_{\mu\nu} dx^{\mu} dx^{\nu} &= \eta_{\mu'\nu'} dx^{\mu'} dx^{\nu'} \\ &= \eta_{\mu'\nu'} \left(\Lambda_{\sigma}^{\mu'} dx^{\sigma} \right) \left(\Lambda_{\rho}^{\nu'} dx^{\rho} \right) \\ &= \Lambda_{\sigma}^{\mu'} \Lambda_{\rho}^{\nu'} \eta_{\mu'\nu'} dx^{\sigma} dx^{\rho} \end{aligned}$$

Let $\sigma \rightarrow \mu$, $\rho \rightarrow \nu$, $\mu' \rightarrow \alpha'$, and $\nu' \rightarrow \beta'$ (since these are just indices), we get:

$$\boxed{\eta_{\mu\nu} = \Lambda_{\sigma}^{\alpha'} \Lambda_{\rho}^{\beta'} \eta_{\alpha'\beta'}}$$

So, we conclude that in Poincaré transformations, the metric tensor transforms according to the above rule. We make two observations, as a consequence of the above fact and from the subsection on Lorentz transformations:

- $\eta_{\mu\nu}$ is a tensor, since it transforms correctly under general coordinate transformations (Poincaré).
- $\eta_{\mu\nu}$ is unchanged (preserved) under Lorentz transformations.

As a summary, we have this little table:

Contravariant vectors: $\lambda^{\mu'} = \Lambda_{\nu}^{\mu'} \lambda^{\nu}$
Covariant vectors: $\lambda_{\mu'} = \Lambda_{\mu'}^{\nu} \lambda_{\nu}$
Tensors: $\tau^{\mu'\nu'}_{\sigma'} = \Lambda_{\alpha}^{\mu'} \Lambda_{\beta}^{\nu'} \Lambda_{\sigma'}^{\gamma} \tau^{\alpha\beta}_{\gamma}$
Scalars: Invariant

Example 4.2. Show that inner products of vectors are scalars.

We actually have done an identical problem before. Consider two vectors $a^{\mu'}$ and $b^{\mu'}$. Their inner product is given by:

$$\begin{aligned} a^{\mu'} b_{\mu'} &= \Lambda_{\nu}^{\mu'} a^{\nu} \Lambda_{\mu'}^{\sigma} b_{\sigma} \\ &= \Lambda_{\nu}^{\mu'} \Lambda_{\mu'}^{\sigma} a^{\nu} b_{\sigma} \\ &= \delta_{\nu}^{\sigma} a^{\nu} b_{\sigma} \\ &= a^{\nu} b_{\nu}. \end{aligned}$$

This shows that the norm of every 4-vector is invariant:

$$\lambda \cdot \lambda = \lambda^{\mu} \lambda_{\mu} = \lambda^{\mu'} \lambda_{\mu'}.$$

Notice that the **sign** of a vector norm squared is also invariant. Among other quantities, this fact has some profound physical consequences. Recall that, in Minkowski spacetime, the norm squared of a vector can be either negative, positive, or zero. We recall and rethink the following facts from introductory special relativity:

- If $\lambda^2 > 0$, then λ^μ is **time-like**. This means there exists a reference frame where $\lambda^{\mu'} = (\lambda^0, 0, 0, 0)$, i.e., there exists a reference frame where the “event” described is at the spatial origin. Recall that if λ^μ is a vector between two events on the Minkowski spacetime diagram, then **time-likeness** implies the existence of a frame where the events occur **at the same location**.
- If $\lambda^2 < 0$, then λ^μ is **space-like**. This means there exists a reference frame where $\lambda^{\mu'} = (0, \lambda^1, 0, 0)$, i.e., there exists a reference frame where the “event” described is at the time origin. Recall that if λ^μ is a vector between two events on the Minkowski spacetime diagram, then **space-likeness** implies the existence of a frame where the events are **simultaneous**.
- If $\lambda^2 = 0$, then λ^μ is **light-like**. We say that λ^μ is a **null vector**, i.e., we can always find a frame where $\lambda^\mu = (\lambda^0, \lambda^0, 0, 0)$, or $(\lambda^0, 0, \lambda^0, 0)$, etc. More generally, $\lambda^\mu = (\lambda^0, \vec{\lambda})$ with $||\vec{\lambda}|| = \lambda^0$.

Example 4.3. Is $X^\mu = (ct, x, y, z)$ a contravariant vector under Poincaré transformations?

In order for $X^\mu = (ct, x, y, z)$ to be a contravariant vector under Poincaré transformation, the following needs to hold:

$$X^{\mu'} = \Lambda_{\nu}^{\mu'} X^{\nu}.$$

Now, the Poincaré transformation is given by:

$$X^{\mu'} = \Lambda_{\nu}^{\mu'} X^{\nu} + a^{\mu'}.$$

for any $a^{\mu'}$. We immediately realize that X^μ is not a vector if $a^{\mu'} \neq 0$, because for both equalities to hold, it is required that $a^{\mu'} = 0$, in which case, the transformation is reduced to a very special case (Lorentz transformation), rather than any general Poincaré transformation where $a^{\mu'}$ can be non-zero.

Example 4.4. Is $dX^\mu = (c dt, dx, dy, dz)$ a contravariant vector under Poincaré transformations?

Recall the form of Poincaré transformations:

$$X^{\mu'} = \Lambda_{\nu}^{\mu'} X^{\nu} + a^{\mu'}.$$

Computing the differentials, we obtain:

$$dX^{\mu'} = \Lambda_{\nu}^{\mu'} dX^{\nu} + 0 = \Lambda_{\nu}^{\mu'} dX^{\nu}.$$

Since dX^{μ} transform correctly, dX^{μ} is a vector under Poincaré transformations.

Example 4.5. Here is an interesting follow-up to the previous example. Suppose that we take $\partial/\partial X^{\mu}$ of a scalar field ϕ . Will the end result be a vector? In other words, is $\partial\phi/\partial X^{\mu}$ a vector? If so, what type of vector is this?

Without much thought, we apply the chain rule on $\phi = \phi(X^{\nu}(X^{\mu'}))$ to obtain the following:

$$\frac{\partial\phi}{\partial X^{\mu'}} = \frac{\partial\phi}{\partial X^{\nu}} \frac{\partial X^{\nu}}{\partial X^{\mu'}} = \Lambda_{\mu'}^{\nu} \frac{\partial\phi}{\partial X^{\nu}}.$$

We observe that $\partial\phi/\partial X^{\mu}$ transforms correctly under Poincaré transformations. Therefore, $\partial\phi/\partial X^{\mu}$ is a vector. Specifically, $\partial\phi/\partial X^{\mu}$ is a **covariant vector**. We can tell from two ways: (i) because the upper indices cancel out, and (ii) by the Λ term in the transformation rule. For covariant transformations from “unprimed” to “primed,” the Λ term has an unprimed upper index and a primed lower index.

The previous example is interesting enough that we should make some definitions, for the sake of simplicity and bookkeeping. For the partial derivative operator with respect to a coordinate, we define

$$\partial_{\mu} = \frac{\partial}{\partial X^{\mu}}$$

This simplifies our notation for a partial derivative of a scalar field:

$$\partial_{\mu}\phi = \frac{\partial\phi}{\partial X^{\mu}}.$$

With this definition, we can construct a vector whose components are the partial derivatives of a scalar field: the **gradient**:

$$\vec{\nabla} = \partial_i = (\partial_1, \partial_2, \partial_3)$$

Of course, we can define a “generalized gradient” operator as:

$$\partial_{\mu} = (\partial_0, \partial_i) = (\partial_0, \vec{\nabla})$$

These definitions will come in handy when we revisit the Maxwell’s equations.

Now that we have a “covariantly” operator, it makes sense to define a “contravariantly” operator ∂^i . In Minkowski spacetime with Cartesian coordinates, we define a lower coordinate:

$$X_\mu = \eta_{\mu\nu} X^\nu.$$

Let

$$\partial^\mu = \frac{\partial}{\partial X_\mu}$$

From

$$X^\mu = \eta^{\mu\nu} X_\nu,$$

we get

$$\frac{\partial X^\mu}{\partial X_\nu} = \eta^{\mu\nu},$$

since $\eta^{\mu\nu}$ is constant. This implies

$$\partial^\mu = \frac{\partial}{\partial X_\mu} = \frac{\partial X^\nu}{\partial X_\mu} \frac{\partial}{\partial X^\nu} = \eta^{\mu\nu} \partial_\nu.$$

We quickly realize that we can raise the index of ∂_μ to get ∂^μ :

$$\partial^\mu = \eta^{\mu\nu} \partial_\nu$$

So, for a scalar field ϕ :

$$\partial^\mu \phi = \eta^{\mu\nu} \partial_\nu \phi.$$

But note that $\partial^i \neq \vec{\nabla}$. Rather, due to the nature of $\eta_{\mu\nu}$ (or equivalently $\eta^{\mu\nu}$), we get

$$\partial^i = -\partial_i = -\vec{\nabla}.$$

So, we can write

$$\partial^\mu = (\partial^0, \partial^i) = (\partial^0, -\vec{\nabla})$$

4.1.4 Velocity, momentum, and force

In this section we want to look at what velocity, momentum, and force look like as 4-vectors and how they transform. To be precise, we want to know in which form do these vectors transform correctly under general Poincaré transformations or specifically in Lorentz transformations. Consider a position vector $X^\mu = (ct, \vec{X})$

in Minkowski spacetime, with metric tensor $\eta_{\mu\nu}$.

Let us conjecture that “velocity” is given by:

$$\frac{dX^{\mu'}}{dt} = \frac{d}{dt} \Lambda_{\nu}^{\mu'} X^{\nu} + \frac{d}{dt} a^{\mu'} = \Lambda_{\nu}^{\mu'} \frac{dX^{\nu}}{dt}.$$

We want to see whether V^{μ} is a vector (specifically 4-vector). Let us call $dX^{\mu}/dt = V^{\mu}$ and $d\vec{x}/dt = \vec{v}$, we obtain:

$$V^{\mu} = (c, \vec{v}).$$

We give the same definitions for the “primed” frame:

$$V^{\mu'} = \frac{dX^{\mu'}}{dt'} = (c, \vec{v}').$$

Since $t \neq t'$ in general,

$$V^{\mu} = \frac{dX^{\mu}}{dt} \neq \frac{dX^{\mu'}}{dt'} = V^{\mu'},$$

which means that

$$V^{\mu'} \neq \Lambda_{\nu}^{\mu'} V^{\nu}.$$

So, no. V^{μ} is not a vector in Minkowski spacetime. Therefore, this definition of “velocity” is not “good.” However, we can actually find a similar 4-vector that works as velocity, using the **proper time**, τ . Consider objects with non-zero mass and speed $v < c$ (so not photons). In this case:

$$ds^2 = c^2 d\tau^2 = \eta_{\mu\nu} dX^{\mu} dX^{\nu} > 0.$$

If we divide both sides by $d\tau^2$, we get the speed of light squared, which is an invariant:

$$c^2 = \eta_{\mu\nu} \frac{dX^{\mu}}{d\tau} \frac{dX^{\nu}}{d\tau}.$$

Let us define the **world velocity**:

$$u^{\mu} = \frac{dX^{\mu}}{d\tau}$$

We see immediately that u^{μ} transforms correctly. Applying the chain rule:

$$u^{\mu'} = \frac{dX^{\mu'}}{d\tau} = \frac{dX^{\mu'}}{dX^{\nu}} \frac{dX^{\nu}}{d\tau} = \Lambda_{\nu}^{\mu'} u^{\nu}.$$

This shows that u^{μ} , as expected by the index placement, is a contravariant 4-vector under general Poincaré transformations. We also see that, by definition,

$$u^{\mu} u_{\mu} = \eta_{\mu\nu} \frac{dX^{\mu}}{d\tau} \frac{dX^{\nu}}{d\tau} = c^2,$$

which is invariant. Next, we can relate u^μ and V^μ by $c^2 d\tau^2 = c^2 dt^2 - d\vec{x}^2$:

$$\frac{d\tau^2}{dt^2} = 1 - \frac{1}{c^2} \frac{d\vec{x}^2}{dt^2} = 1 - \frac{v^2}{c^2} = \frac{1}{\gamma^2}.$$

We have just derived **time-dilation** (for massive objects of course):

$$\boxed{\frac{dt}{d\tau} = \gamma}$$

This implies:

$$\boxed{u^\mu = \gamma V^\mu = \gamma(c, \vec{v}) = (\gamma c, \gamma \vec{v})}$$

In the object's rest frame, $\vec{v} = 0$, $\gamma = 1$. This gives $u^\mu = (c, 0, 0, 0)$. We can interpret this as "the object travels at speed c in the time direction when in its rest frame."

Moving on to **momentum**. We simply define 4-momentum based on 4-velocity:

$$\boxed{p^\mu = m u^\mu}$$

We can quickly see that

$$\boxed{p^\mu = \gamma \left(\frac{mc^2}{c}, m\vec{v} \right) = \left(\frac{\gamma mc^2}{c}, \gamma m\vec{v} \right) = \left(\frac{E}{c}, \vec{p} \right)}$$

Next, we shall look at what the norm square (invariant) of 4-momentum is. It is easy to show that:

$$p^\mu p_\mu = m^2 u^\mu u_\mu = m^2 c^2.$$

But also, by definition:

$$p^\mu p_\mu = \eta_{\mu\nu} dp^\mu d^\nu = \frac{E^2}{c^2} - \vec{p}^2 = \frac{E^2}{c^2} - p^2.$$

The above two equalities give us the full energy-momentum equivalence

$$\boxed{E^2 = c^2 p^2 + m^2 c^4}$$

But what about massless particles (like photons)? For massless particles in Minkowski spacetime, $v = c$, and hence the proper time τ is not defined. This also means the definition $u^\mu = dX^\mu/d\tau$ is no longer defined. However, the line element ds^2 still is:

$$\begin{aligned} ds^2 &= c^2 dt^2 - d\vec{x}^2 \\ &= c^2 dt^2 \left(1 - \frac{1}{c^2} \frac{d\vec{x}^2}{dt^2} \right) \\ &= c^2 dt^2 \left(1 - \frac{c^2}{c^2} \right) \\ &= 0. \end{aligned}$$

It is indeed true that for light

$$\boxed{ds^2 = 0}$$

by the construction of Minkowski spacetime. Recall that photon vectors are “light-like,” as we often said special relativity. Here, we will be using a somewhat different terminology. We will refer to photon’s trajectories as **null** in flat Minkowski spacetime. This can be readily shown, as follows

Now, just because the proper time τ is not defined does not mean we cannot describe the trajectory of light. Consider a parameterization $X^\mu = X^\mu(\sigma)$, where σ is simply some parameter. Let us define $u^\mu = \frac{\partial X^\mu}{\partial \sigma}$, then, since $ds^2 = 0$

$$u^\mu u_\mu = \eta_{\mu\nu} \frac{\partial X^\mu}{\partial \sigma} \frac{\partial X^\nu}{\partial \sigma} = \frac{ds^2}{d\sigma^2} = 0,$$

i.e., u^μ is light-like. We can also talk about energy and momentum of light. Let p^μ be the 4-vector momentum of light, then

$$p^\mu = \left(\frac{E}{c}, \vec{p} \right) = (p^0, \vec{p}).$$

But recall that for light, $E = h\nu$ and $||\vec{p}|| = h/\lambda$, i.e., $E = c||\vec{p}||$. This gives

$$p^\mu p_\mu = \eta_{\mu\nu} dp^\mu dp^\nu = \frac{E^2}{c^2} - ||\vec{p}||^2 = \frac{E^2}{c^2} - \frac{E^2}{c^2} = 0,$$

which should make sense, because light 4-momentum is a light-like vector. We shall also see that the wave 4-vector, $k^\mu = (k^0, \vec{k})$, defined as $p^\mu = \hbar k^\mu$, is also a light-like vector. We first notice that

$$k^0 = \frac{p^0}{\hbar} = \frac{h}{\lambda} \frac{1}{\hbar} = \frac{2\pi}{\lambda} = ||\vec{k}||.$$

This implies

$$k^\mu k_\mu = (k^0)^2 - ||\vec{k}||^2 = ||\vec{k}||^2 - ||\vec{k}||^2 = 0.$$

This is quite obvious, because $k^\mu \propto p^\mu$.

Example 4.6. Find the wavelength λ for light emitted from a source (with emitted wavelength λ_0) that is receding at speed v .

Let the wave vector in the moving frame (S') be $k^{\mu'} = (k^{0'}, \vec{k}')$. Since the source is in (S'),

$$k^{\mu'} = (k^{0'}, \vec{k}') = \left(\frac{2\pi}{\lambda_0}, \frac{-2\pi}{\lambda_0}, 0, 0 \right).$$

In the stationary frame (S),

$$k^\mu = (k^0, \vec{k}) = \left(\frac{2\pi}{\lambda}, \frac{-2\pi}{\lambda}, 0, 0 \right).$$

Now, since we want to find a quantity in the stationary frame, we want to apply the inverse transformation $\Lambda_{\nu'}^\mu$ to $k^{\nu'}$ in (S') to get back k^μ in (S):

$$k^{\mu'} = \Lambda_{\nu'}^\mu k^\nu.$$

As a matrix equation:

$$2\pi \begin{pmatrix} \lambda^{-1} \\ -\lambda^{-1} \\ 0 \\ 0 \end{pmatrix} = 2\pi \begin{pmatrix} \gamma & \beta\gamma & 0 & 0 \\ \beta\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \lambda_0^{-1} \\ -\lambda_0^{-1} \\ 0 \\ 0 \end{pmatrix}$$

It is not difficult to see that

$$\frac{2\pi}{\lambda} = \Lambda_{\nu'}^0 k^{\nu'} = \frac{2\pi\gamma}{\lambda_0} - \gamma\beta \frac{2\pi}{\lambda_0} = \frac{2\pi}{\lambda_0} \gamma(1 - \beta).$$

Therefore,

$$\frac{1}{\lambda} = \frac{\gamma}{\lambda_0} (1 - \beta) = \frac{1}{\lambda_0} \sqrt{\frac{1 - \beta}{1 + \beta}}.$$

We obtain nothing but the **relativistic Doppler shift** formula for a receding light source.

$$\lambda = \lambda_0 \sqrt{\frac{1 + \beta}{1 - \beta}}$$

We notice that because $\lambda > \lambda_0$, the observed light is **red-shifted**. Of course, if we have an approach source, then $v \rightarrow -v$, which gives **blue-shifted** observed light:

$$\lambda = \lambda_0 \sqrt{\frac{1 - \beta}{1 + \beta}}$$

Keep in mind that these are Doppler shifts due to relative motion. Later on in this text, we will discuss **gravitation spectral shifts** and **Cosmological red-shift**. The first is due the curvature of spacetime, whereas the latter is due to the expansion of space (the universe).

Just as 4-vector momentum is defined on 4-vector velocity (or the world velocity) as its scalar multiple, **4-vector force**, f^μ , for massive objects, is defined on 4-vector momentum as

$$f^\mu = \frac{dp^\mu}{d\tau}$$

With $p^\mu = mu^\mu$, we get the relativistic version of Newton's 2nd law:

$$f^\mu = m \frac{d^2 X^\mu}{d\tau^2}.$$

Next, we want to relate the 4-force with our much-familiar 3-force, \vec{F} . We can start with $p^\mu = (E/c, \vec{p})$. Applying the chain rule to the 4-momentum, and the fact that $dt/d\tau = \gamma$,

$$\frac{dp^\mu}{d\tau} = \frac{dt}{d\tau} \frac{dp^\mu}{dt} = \gamma \frac{dp^\mu}{dt} = \gamma \left(\frac{1}{c} \frac{dE}{dt}, \frac{d\vec{p}}{dt} \right).$$

Noticing that $dE/dt = \vec{F} \cdot \vec{v}$ is the definition of “power,” we get:

$$f^\mu = \gamma \left(\frac{1}{c} \vec{F} \cdot \vec{v}, \vec{F} \right)$$

Example 4.7. Show that $u^\mu f_\mu = 0$.

Without loss of generality (recall: scalars are invariant), we can assume a 1-dimensional case where

$$f^\mu = \left(\frac{\gamma v F}{c}, \gamma F, 0, 0 \right)$$

and

$$u^\mu = (\gamma c, \gamma v, 0, 0).$$

We can first show that $u^\mu f_\mu = 0$, by a mechanical computation:

$$u^\mu f_\mu = \eta_{\mu\nu} u^\mu f^\nu = \gamma^2 v F - \gamma^2 v F = 0.$$

We can also solve this problem geometrically. In Minkowski's spacetime diagram, u^μ and f^μ are “orthogonal,” not in the sense that they are perpendicular (note that this is true only in Euclidean geometry), but rather in the sense that there exists a frame (S') such that $u^{\mu'}$ is “parallel” to the ct' axis and $f^{\mu'}$ is parallel to the x' axis, i.e., they are orthogonal in (S'). But because their dot product is invariant under Lorentz transformations, $u^\mu f_\mu = u^{\mu'} f_{\mu'} = 0$.

insert figure 13 here!!!

4.2 Relativistic Electrodynamics

Previously we have obtained Maxwell's equations in differential form, from the integral form, with $\vec{E} = (E^1, E^2, E^3) = E^i$ and $\vec{B} = (B^1, B^2, B^3) = B^j$. ρ is the charge density and $\vec{J} = (J^1, J^2, J^3) = J^k$ is the current density.

$$\text{Gauss' law: } \oint \vec{E} \cdot d\vec{a} = \frac{q}{\epsilon_0}$$

$$\text{No magnetic monopole: } \oint \vec{B} \cdot d\vec{a} = 0$$

$$\text{Faraday's law: } \oint \vec{E} \cdot d\vec{s} = -\frac{d}{dt}\Phi_B = \frac{\partial}{\partial t} \int_A \vec{B} \cdot d\vec{a}$$

$$\text{Ampere-Maxwell's law: } \oint \vec{B} \cdot d\vec{s} = \mu_0 I + \mu_0 \epsilon_0 \frac{\partial}{\partial t} \int_A \vec{E} \cdot d\vec{a}$$

Note that \vec{E} and \vec{B} are 3-vectors. Together, their six components mix under Lorentz transformations, i.e., \vec{E} and \vec{B} join relativistically. We have been working in 4-dimensional flat Minkowski spacetime, and we would want to rewrite Maxwell's equations in this space. How are we going to accomplish this? Unfortunately, the full answer is beyond the scope of this text. However, let it be given a tensor, which is a combination of \vec{E} and \vec{B} in spacetime. We have the following definition:

Definition 4.1. The electromagnetic field strength tensor $F^{\mu\nu}$ is given by:

$$[F^{\mu\nu}] = \begin{pmatrix} 0 & E^1/c & E^2/c & E^3/c \\ -E^1/c & 0 & B^3 & -B^2 \\ -E^2/c & -B^3 & 0 & B^1 \\ -E^3/c & B^2 & -B^1 & 0 \end{pmatrix}$$

Now, applying index-lowering operations twice gives

$$F_{\mu\nu} = \eta_{\mu\alpha} \eta_{\nu\beta} F^{\alpha\beta}$$

As matrices,

$$[F_{\mu\nu}] = [\eta_{\mu\alpha}] [F^{\alpha\beta}] [\eta_{\nu\beta}]^\top$$

This gives

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

Next, we can form vectors out of ρ and \vec{J} , the **4-vector current density**:

$$j^\mu = (\rho c, \vec{J})$$

The Maxwell equations, in index notation become:

$$\begin{aligned} \partial_\nu F^{\mu\nu} &= \mu_0 j^\mu \\ \partial_\sigma F_{\mu\nu} + \partial_\mu F_{\nu\sigma} + \partial_\nu F_{\sigma\mu} &= 0 \end{aligned}$$

We shall see readily how the equations above manifest in differential form. First, we look at

$$\partial_\nu F^{\mu\nu} = \mu_0 j^\mu.$$

For $\mu = 0$,

$$\begin{aligned} \partial_\nu F^{0\nu} &= \mu_0 j^0 = \mu_0 \rho c \\ &= \partial_0 F^{00} + \partial_1 F^{01} + \partial_2 F^{02} + \partial_3 F^{03} \\ &= 0 + \frac{1}{c} \partial_i E^i \\ &= \frac{1}{c} \vec{\nabla} \cdot \vec{E}. \end{aligned}$$

But because $\mu_0 \epsilon_0 = c^{-2}$,

$$\vec{\nabla} \cdot \vec{E} = \rho c^2 \mu_0 = \frac{\rho}{\epsilon_0},$$

which is Gauss' law in differential form. Next, we let $\mu = k = \{1, 2, 3\}$. So,

$$\begin{aligned} \partial_\nu F^{k\nu} &= \mu_0 j^k = \mu_0 J^k \\ &= \partial_0 F^{k0} + \partial_i F^{ki} \\ &= \frac{1}{c} \frac{\partial}{\partial t} \left(\frac{-E^k}{c} \right) + \partial_i F^{ki} \\ &= -\frac{1}{c^2} \frac{\partial E^k}{\partial t} + \partial_i F^{ki}. \end{aligned}$$

To find what $\partial_i F^{ki}$ is, we can look at the case where $k = 1$:

$$\begin{aligned} \partial_i F^{1i} &= \partial_1 F^{11} + \partial_2 F^{12} + \partial_3 F^{13} \\ &= 0 + \partial_2 B^3 \partial_3 (-B^2) \\ &= \left(\vec{\nabla} \times \vec{B} \right)^{-1}. \end{aligned}$$

So, $\partial_i F^{ki}$ is the **curl**!

$$\partial_i F^{ki} = \left(\vec{\nabla} \times \vec{B} \right)^k$$

Combining this with what we have found earlier:

$$-\frac{1}{c^2} \frac{\partial E^k}{\partial t} + \left(\vec{\nabla} \times \vec{B} \right)^k = \mu_0 J^k.$$

This is nothing but Ampere-Maxwell's law:

$$\boxed{\vec{\nabla} \times \vec{B} = \mu_0 \vec{J} + \mu_0 \epsilon_0 \frac{\partial \vec{E}}{\partial t}}$$

Similarly, we can look at the other equation:

$$\partial_\sigma F_{\mu\nu} + \partial_\mu F_{\nu\sigma} + \partial_\nu F_{\sigma\mu} = 0.$$

We can show that for various values of σ, ν, μ , we get Faraday's law and the "law of no magnetic monopole."

Example 4.8. Show that, for $\mu = 0, \nu = 1, \sigma = 2$, we get

$$\left(\vec{\nabla} \times \vec{E} \right)^3 = - \left(\frac{\partial \vec{B}}{\partial t} \right)^3$$

This is just a simple plug-and-chug problem:

$$\begin{aligned} \partial_2 F_{01} + \partial_0 F_{12} + \partial_1 F_{20} &= \frac{\partial}{\partial y} \left(-\frac{E^1}{c} \right) + \frac{1}{c} \frac{\partial B^3}{\partial t} + \frac{\partial}{\partial x} \left(\frac{E^2}{c} \right) \\ &= \left(\vec{\nabla} \times \vec{E} \right)^3 + \left(\frac{\partial \vec{B}}{\partial t} \right)^3 \\ &= 0. \end{aligned}$$

So,

$$\left(\vec{\nabla} \times \vec{E} \right)^3 = - \left(\frac{\partial \vec{B}}{\partial t} \right)^3$$

This result can of course be generalized to give

$$\left(\vec{\nabla} \times \vec{E} \right)^j = - \left(\frac{\partial \vec{B}}{\partial t} \right)^j.$$

Or equivalently, Faraday's law.

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

To show the “no magnetic monopole” law, we can limit our use to only the non-zero indices, so that what comes out of the cyclic identity is the divergence of \vec{B} , or $\partial_i B^i = 0$. This is left as an exercise for the reader.

So, to summarize, in special relativity, all physical properties are some sort of tensors with (invariant) scalars: m, τ, ds^2, c . We also have the special vectors (which are a type of tensors): the world velocity u^μ , 4-momentum p^μ , and 4-force f^μ . Notable tensors include the Minkowski metric $\eta_{\mu\nu} = \eta^{\mu\nu}$, and the electromagnetic field strength tensor $F^{\mu\nu}$, all of which transform correctly under Lorentz transformations.

5 Curved spaces

Recall that equivalence principle leads us towards the idea of curved spacetime. In general relativity, gravity is not a force. Instead, massive objects curve or wrap spacetime around them. Light travels as a free particle along a “geodesic” through curved spacetime. One of the questions we will eventually answer in this section is: “How to find the equation for geodesics?” To answer this question, we will have to learn how to describe curved spaces and spacetime directly. We will derive the **geodesic equation**:

$$\frac{d^2 X^\mu}{d\tau^2} + \Gamma^\mu_{\nu\sigma} \frac{dX^\nu}{d\tau} \frac{dX^\sigma}{d\tau} = 0.$$

We will also see how the metric tensor $g_{\mu\nu}$ and the **Christoffel symbol** $\Gamma^\mu_{\nu\sigma}$, and the **Riemann curvature tensor** $R^\lambda_{\mu\nu\sigma}$ are related. We, of course, will explore what these mathematical objects are. Then (finally), we will look at the Einstein equations that will let us solve the metric tensor $g_{\mu\nu}$ for a given distribution of matter (mass and energy).

5.1 Geodesics

In 3-dimensional flat space, we can think of a **geodesic** as the shortest distance between two points, i.e., a straight line, or the path of a free particle. It is very important to remember that **free particles follow geodesics**, as this is true in all spaces, while geodesics can be curves.

For general n -dimensional spaces, geodesics are no longer shortest distances, as we will see in this section. Rather, geodesics are paths of free particles.

Example 5.1. Show that a geodesic in Minkowski spacetime is not the shortest distance.

Consider a free particle, $f^\mu = 0$. This implies

$$\frac{\partial^2 X^\mu}{\partial \tau^2} = 0,$$

which means there exists a geodesic X^μ in spacetime such that $X^\mu = X^\mu(\tau)$ is a **straight line in spacetime** (the equation above, in terms of vector calculus, means the curve has no curvature). However, we can show now that this X^μ does not give the shortest distance in spacetime. Consider

$$ds^2 = \eta_{\mu\nu} dX^\mu dX^\nu.$$

For moving, massive particles, X^μ is time-like:

$$ds^2 = c^2 d\tau^2 > 0$$

Now, consider two points A and B on a Minkowski's spacetime diagram and consider two possible paths from A to B .

insert figure 14 here!!!

The first path (vertical) indicates that the particle is at rest in spacetime (so $f^\mu = 0$). The second path is that of a moving particle. Now, because they depart and arrive at the same locations in spacetime, $\Delta\tau = \Delta t$. Next, we calculate the “length” of the first path as $c\Delta\tau$, as evidently shown on the spacetime diagram. What about the length of the second path?

$$c\Delta\tau' = 2\sqrt{\left(\frac{1}{2}c\Delta t\right)^2 - (\Delta x)^2} < c\Delta\tau^2.$$

This implies

$$\Delta\tau' < \Delta\tau,$$

i.e., the geodesic maximizes the proper time τ , or equivalently, the “length” of the geodesic in this case is the largest of all possible “lengths.”

5.2 2-dimensional curved spaces

According to general relativity, we live in a 4-dimensional spacetime. However, this is hard to visualize. To start off more simply, we can look at 2-dimensional spaces that we can embed in 3-dimensional spaces.

First off, 2-dimensional spaces can be embedded in 3-dimensional spaces. For example, the xy -plane can be embedded into 3-space simply by letting $z = 0$ in Cartesian coordinates. Just like how 1-dimensional curves in 3-dimensions can be described by a parameterization $\vec{r} = \vec{r}(t) = (x(t), y(t), z(t))$, 2-dimensional surfaces can be parameterized with 2 parameters, say u and v . For example, a surface \vec{S} in \mathbb{R}^3 can have the following form:

$$\vec{S} = \vec{S}(u, v) = (x(u, v), y(u, v), z(u, v)).$$

Example 5.2. Parameterize the surface of a sphere of radius R .

We know from spherical coordinates the conversion formulas:

$$\begin{aligned} x &= R \sin \theta \cos \phi \\ y &= R \sin \theta \sin \phi \\ z &= R \cos \theta. \end{aligned}$$

So the parameterization is simply:

$$\vec{S}(\theta, \phi) = (R \sin \theta \cos \phi, R \sin \theta \sin \phi, R \cos \theta),$$

where $0 \leq \phi \leq 2\pi$ and $0 \leq \theta \leq \pi$. We can give this parameterization to Mathematica, under “ParametricPlot3D” to get a sphere. Let $R = 2$.

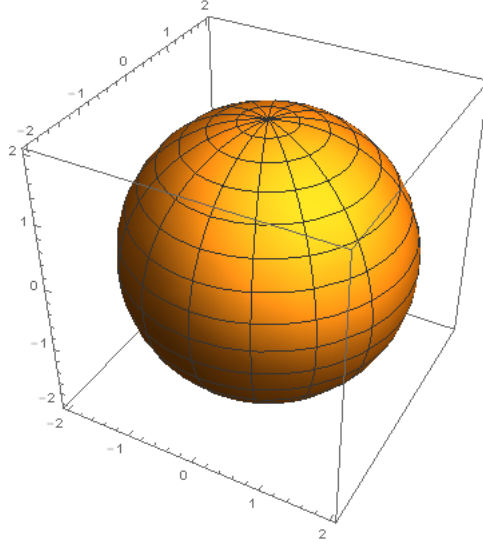


Figure 8: Surface of a sphere of radius $R = 2$

But we in fact don't need an extra dimension to describe a curved space. An obvious example is that we don't need to leave the surface of Earth to know that it is curved. With just two coordinates u and v in a 2-dimensional space, we can mathematically show whether the space is curved or not. The first step is to generate tangent vectors:

$$\vec{e}_i = \frac{\partial \vec{r}}{\partial u^i}.$$

For the case of a 2-dimensional surface parameterized by u and v , we obtain two tangent vectors:

$$\vec{e}_u = \frac{\partial \vec{S}}{\partial u}, \text{ and } \vec{e}_v = \frac{\partial \vec{S}}{\partial v}.$$

Note that these tangent vectors are not in the surface (this is also true in general n -dimensional cases). Rather, they lie in a tangent space T_p at each point P in the space. Tangent vectors are useful because they give directions along the curve or surface.

Next, we want to look at an infinitesimal displacement $d\vec{S}$. Because $\vec{S} = \vec{S}(u, v)$, we get from the chain rule:

$$d\vec{S} = \frac{\partial \vec{S}}{\partial u} du + \frac{\partial \vec{S}}{\partial v} dv = \vec{e}_u du + \vec{e}_v dv.$$

For the sake of convenience (and further generalizations in the future), we shall switch to index notation. Recall that capital Roman characters are reserved for

2-dimensional manifolds, let $u^A = (u^1, u^2) = (u, v)$, with $A = 1, 2$. This gives

$$d\vec{S} = \vec{e}_A du^A.$$

It follows that the line element ds^2 is

$$ds^2 = d\vec{S} \cdot d\vec{S} = (\vec{e}_A du^A) \cdot (\vec{e}_B du^B) = \vec{e}_A \cdot \vec{e}_B du^A du^B.$$

But this is nothing but the all-familiar:

$$\boxed{ds^2 = g_{AB} du^A du^B}$$

Note that we have just shown the above equation holds not only for flat but also for general curved spaces. With this, we can go on to calculate lengths of curves in general curved 2-dimensional spaces. Let a curve in this 2-dimensional space be $\vec{l}(\sigma) = (u(\sigma), v(\sigma))$. The length L of this curve for $a \leq \sigma \leq b$ is given by:

$$\begin{aligned} L &= \int_a^b ds \\ &= \int_a^b \sqrt{g_{AB} du^A du^B} \\ &= \int_a^b \sqrt{g_{AB} \frac{du^A}{d\sigma} \frac{du^B}{d\sigma}} d\sigma \\ &= \int_a^b \sqrt{g_{AB} \dot{u}^A(\sigma) \dot{u}^B(\sigma)} d\sigma. \end{aligned}$$

We realize that this has the same form as the length-formula we have found earlier. The only difference is that we are in curved spaces. This suggests that we can generalize this formula for any n -dimensional space.

We have defined the natural basis vectors as $\vec{e}_A = \partial\vec{S}/\partial u^A$. What about the dual basis vectors? If we try to define the dual basis vectors as $\vec{e}^A = \vec{\nabla} u^A$, we run into trouble. Recall that we had we previous had three coordinates, and that the gradient of u is a normal vector to the surface of constant u . However, in 2-dimensional spaces constant u are lines, which has infinitely many normal vectors. This simply means we cannot use the gradient operator to find \vec{e}^A . Rather, we will use a more powerful tool, the **inverse metric tensor**, g^{AB} to raise the index of the natural basis vectors \vec{e}_A :

$$\boxed{\vec{e}^A = g^{AB} \vec{e}_B}$$

So, the general procedure is: (i) find the tangent vectors, i.e., the natural basis vectors \vec{e}_A , (ii) generate the metric tensor g_{AB} , (iii) compute the inverse metric tensor g^{AB} , and (iv) find the dual basis vectors \vec{e}^A from raising the index of \vec{e}_A .

Example 5.3. Consider a saddle in 3-dimensional flat space. Using paraboloidal coordinates with constant w , find the metric tensor and inverse metric tensor of this saddle surface.

Recall the conversion formulas for paraboloidal coordinates:

$$x = u + v$$

$$y = u - v$$

$$z = 2uv$$

Let the saddle be

$$\vec{r} = (u + v, u - v, 2uv).$$

We first find the tangent vectors (or the natural basis vectors):

$$\vec{e}_1 = \frac{\partial \vec{r}}{\partial u} = (1, 1, 2v)^\top$$

$$\vec{e}_2 = \frac{\partial \vec{r}}{\partial v} = (1, -1, 2u)^\top.$$

Finding the metric tensor is then purely mechanical:

$$[g_{AB}] = [\vec{e}_A \cdot \vec{e}_B] = \begin{pmatrix} 2 + 4v^2 & 4uv \\ 4uv & 2 + 4u^2 \end{pmatrix},$$

as for finding the inverse metric tensor:

$$[g^{AB}] = [g_{AB}]^{-1} = \frac{1}{2(1 + 2u^2 + 2v^2)} \begin{pmatrix} 1 + 2u^2 & -2uv \\ -2uv & 1 + 2v^2 \end{pmatrix}$$

With the natural basis vectors and the inverse metric tensor, we can easily find the dual basis vectors:

$$\begin{aligned} \vec{e}^1 &= [g^{AB}]\vec{e}_1 = \frac{1}{2(1 + 2u^2 + 2v^2)} \begin{pmatrix} 1 + 2u^2 & -2uv \\ -2uv & 1 + 2v^2 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 2v \end{pmatrix} \\ \vec{e}^2 &= [g^{AB}]\vec{e}_2 = \frac{1}{2(1 + 2u^2 + 2v^2)} \begin{pmatrix} 1 + 2u^2 & -2uv \\ -2uv & 1 + 2v^2 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \\ 2u \end{pmatrix} \end{aligned}$$

Ultimately, though, we will find that basis vectors are not so useful once we already have the metric tensor. So we will not be worrying about them so much moving forward. In the next sections, we will explore the Einstein equations, whose purpose is to find the metric tensor describing a space of a given matter distribution. Once we obtain the metric tensor, we can just mechanically work out other characteristics of the space.

5.3 Manifolds

An arbitrary, curved n -dimensional space is called a manifold. Assuming that we know the metric tensor, we can write the n coordinates associated with the space as

$$x^a = (x^1, x^2, \dots, x^n),$$

for any coordinate system. Next, assume that we have two coordinate systems whose conversion formulae are differential and invertible functions:

$$\begin{aligned} x^{a'} &= x^{a'}(x^b) \\ x^b &= x^b(x^{a'}). \end{aligned}$$

Next, just like before, we define the Jacobian matrix for the transformation:

$$[X_b^{a'}] = \left[\frac{\partial X^{a'}}{\partial X^b} \right]$$

And of course the inverse transformation:

$$[X_{b'}^a] = \left[\frac{\partial X^a}{\partial X^{b'}} \right]$$

These have the usual properties we have seen before:

$$X_b^{a'} X_{c'}^b = \delta_c^a = X_{b'}^a X_c^{b'}$$

Again, just as before, we define vectors, tensors, and scalars by how the transform:

<p>Contravariant vectors: $\lambda^{a'} = X_b^{a'} \lambda^b$</p> <p>Covariant vectors: $\mu_{a'} = X_{a'}^b \mu_b$</p> <p>Tensors: $\tau^{a'b'}_{c'd'} = X_e^{a'} X_f^{b'} X_{c'}^g X_{d'}^h \tau^{ef}_{gh}$</p> <p>Raising/lowering index: $\tau^a_b = g_{bd} \tau^{ad}$</p> <p>Identity: $g^{ab} g_{bc} = \delta_c^a$</p>

In general, the metric tensor need not be **positive definite**, which means that ds^2 can be positive, negative, or zero. However, the **signature** (the number of positives minus the number of negatives) of g_{ab} is a scalar and is -2 for all metric in general relativity. For instance, $\eta_{\mu\nu}$ has signature -2 .

There are classes of manifolds: **Riemannian manifolds**, which have positive definite metric tensors, and **pseudo-Riemannian manifolds**, where metric tensors do not have to be positive definite. Note that spacetime is a pseudo-Riemannian manifold.

To define lengths and distances as real numbers, we need to add the absolute values to our previous definitions:

<p>Infinitesimal distance: $ds = \sqrt{ g_{ab} dx^a dx^b }$</p> <p>Length of a curve: $L = \int_a^b ds = \int_a^b \sqrt{ g_{ab} dx^a dx^b }$</p> <p>Vector norm: $\vec{\lambda} = \sqrt{ \lambda^a \lambda_a }$</p>
--

Note that vectors can be null or non-null. For non-null vectors in general n -dimensional manifolds, we can also define an “angle” θ between two vectors:

$$\cos \theta = \frac{\vec{\lambda} \cdot \vec{\mu}}{||\vec{\lambda}|| ||\vec{\mu}||} = \frac{g_{ab} \lambda^a \lambda^b}{||\vec{\lambda}|| ||\vec{\mu}||}.$$

Beware that this definition works well (intuitively) with positive definite metric tensors. However, for spacetime (whose metric tensor is not positive definite), things can get a little weird.

Example 5.4. Consider space-like vector λ^μ in flat 4-dimensional Minkowski spacetime. Find θ between it and itself.

Without changing the inner product (due to invariance), consider a frame where $\lambda^\mu = (0, \vec{\lambda})$. According to our definition:

$$\cos \theta = \frac{\eta_{\mu\nu} \lambda^\mu \lambda^\nu}{||\lambda^\mu||^2} = \frac{-||\vec{\lambda}||^2}{||\vec{\lambda}||^2} = -1.$$

This implies $\theta = \pi$, so the vector makes an angle of π with itself.

We call the vectors λ^μ and π^μ **orthogonal** if $\lambda^\mu \cdot \pi^\mu = 0$, i.e., there exists a frame where they are orthogonal.

5.4 Tensors on manifolds

5.4.1 Combining tensors

In this section, we will learn the mathematical properties of tensors including special identities and tensor algebra. Familiarizing ourselves with tensor algebra will save us a lot of time and effort when we cover the later topics where we will encounter a few more mathematical objects that are defined on tensors.

Proposition 5.1. Adding tensors of the same type gives tensors.

Proof. Let two general tensors τ^{ab}_c and Θ^{ab}_c be given. Let $\zeta^{a'b'}_{c'} = \tau^{a'b'}_{c'} + \Theta^{a'b'}_{c'}$. We will show that $\zeta^{a'b'}_{c'}$ is a tensor, i.e.:

$$\zeta^{a'b'}_{c'} = X^a_d X^b_e X^f_{c'} \zeta^{de}_f.$$

The proof is a straightforward application of the transformation rules for tensors:

$$\begin{aligned} \zeta^{a'b'}_{c'} &= \tau^{a'b'}_{c'} + \Theta^{a'b'}_{c'} \\ &= X^a_d X^b_e X^f_{c'} \tau^{de}_f + X^a_d X^b_e X^f_{c'} \Theta^{de}_f \\ &= X^a_d X^b_e X^f_{c'} (\tau^{de}_f + \Theta^{de}_f) \\ &= X^a_d X^b_e X^f_{c'} \zeta^{de}_f. \end{aligned}$$

Since ζ^{de}_f transforms correctly, it is a tensor. \square

Proposition 5.2. Multiplying a tensor by a scalar gives a tensor

Proof. The proof to this proposition is trivial. Suppose $\sigma^a_b = \alpha \tau^a_b$, where τ^a_b is a tensor. We want to show that σ^a_b is also a tensor, i.e.,

$$\sigma^{a'}_{b'} = X^{a'}_c X^d_{b'} \sigma^c_d.$$

Once again, a straightforward application of the transformation rule applies:

$$\begin{aligned} \sigma^{a'}_{b'} &= \alpha \tau^{a'}_{b'} \\ &= \alpha X^{a'}_c X^d_{b'} \tau^c_d \\ &= X^{a'}_c X^d_{b'} \sigma^c_d. \end{aligned}$$

\square

So far, we have seen that properties of tensors are similar to those of vectors. Specifically, we see that tensor spaces, similar to vector spaces, are also closed under addition and scalar multiplication. But there are more properties that tensor spaces have but vector spaces do not.

Proposition 5.3. Products of tensors are tensors.

Proof. Suppose that $\sigma^{ab}_c = \lambda^a \tau^b_c$. We want to show:

$$\sigma^{a'b'}_{c'} = X^{a'}_d X^{b'}_e X^f_{c'} \sigma^{de}_f$$

Once again...

$$\begin{aligned} \sigma^{a'b'}_{c'} &= \lambda^{a'} \tau^{b'}_{c'} \\ &= \left(X^{a'}_d \lambda^d \right) \left(X^{b'}_e X^f_{c'} \tau^e_f \right) \\ &= X^{a'}_d X^{b'}_e X^f_{c'} \lambda^d \tau^e_f \\ &= X^{a'}_d X^{b'}_e X^f_{c'} \sigma^{de}_f \end{aligned}$$

\square

Proposition 5.4. Contracting a tensor of type (r, s) gives a tensor of type $(r - 1, s - 1)$.

Proof. Suppose τ_{cd}^{ab} is a $(2, 2)$ tensor. Let us call another tensor $\sigma_b^a = \tau_{cb}^{ab}$, which is a $(1, 1)$ tensor. We want to show that the contraction σ_b^a is also a tensor:

$$\sigma_{b'}^{a'} = X_d^{a'} X_b^g \sigma_g^d.$$

Again, applying the transformation rules:

$$\begin{aligned} \sigma_{b'}^{a'} &= \tau_{c'b'}^{a'c'} \\ &= X_d^{a'} X_{e'}^{c'} X_{c'}^f X_{b'}^g \tau_{fg}^{dc} \\ &= X_d^{a'} X_{b'}^g \delta_f^e \tau_{eg}^{de} \\ &= X_d^{a'} X_{b'}^g \tau_{eg}^{de} \\ &= X_d^{a'} X_{b'}^g \sigma_g^d. \end{aligned}$$

□

Note that we have already seen this property already in the index lowering/raising operations. For example:

$$\lambda_a = g_{ab} \lambda^b.$$

As a consequence of the above proposition, we can let σ^{ab}_c be

$$\sigma^{ab}_c = \tau^{abe} \mu_e g_{cf} \lambda^f,$$

and know immediately that it is a tensor (since the indices “add up to” or “contract to” to upper ab and lower c).

Proposition 5.5 (Diving: Quotient Theorem). Suppose $\tau_{bc}^a \lambda^c$ transforms as a tensor for all λ^c . The **quotient theorem** says τ_{bc}^a is a tensor.

Proof. We want to show that

$$\tau_{b'g'}^{a'} = X_d^{a'} X_{b'}^e X_{g'}^f \tau_{ef}^d.$$

We first know that the given quantity transforms correctly as a tensor. We also know that $\lambda^{c'} = X_f^{c'} \lambda^f$. It follows that

$$\begin{aligned} \tau_{b'c'}^{a'} \left(X_f^{c'} \lambda^f \right) - X_d^{a'} X_{b'}^e \tau_{ef}^d \lambda^f &= 0 \\ \tau_{b'c'}^{a'} X_f^{c'} - X_d^{a'} X_{b'}^e \tau_{ef}^d &= 0 \\ \tau_{b'c'}^{a'} X_f^{c'} &= X_d^{a'} X_{b'}^e \tau_{ef}^d \\ \tau_{b'c'}^{a'} \delta_{g'}^{c'} &= X_d^{a'} X_{b'}^e X_{g'}^f \tau_{ef}^d \\ \tau_{b'g'}^{a'} &= X_d^{a'} X_{b'}^e X_{g'}^f \tau_{ef}^d. \end{aligned}$$

Therefore, τ_{bc}^a is a tensor. □

5.4.2 Special tensors

Symmetric tensors. Let a two-indexed tensor τ_{ab} be given. It is symmetric if

$$\tau_{ab} = \tau_{ba}.$$

Note that a straightforward application of the transformation rule can show that if a tensor is symmetric in one frame (S), then it is also symmetric in a different frame (S').

Anti-symmetric tensors. Let a 2-indexed tensor τ^{ab} be given. We say that τ_{ab} is anti-symmetric if

$$\tau_{ab} = -\tau^{ba}.$$

Again, an application of the transformation rule can show that if a tensor is anti-symmetric in one frame, it is anti-symmetric in all frames.

Kronecker delta. Recall that the Kronecker delta (a type $(1, 1)$ tensor) is defined as

$$\delta_b^a = \begin{cases} 1, a = b \\ 0, a \neq b. \end{cases}$$

As we might have seen before, the Kronecker delta is frame-independent:

$$\delta_b^a = X_e^{a'} X_{b'}^f \delta_f^e = X_e^{a'} X_{b'}^e = \delta_{b'}^{a'}.$$

Note: For most tensors (with any arbitrary number of indices), the ordering of indices matter, i.e., $\tau^{abe} \neq \tau^{aeb}$ in general. Therefore, if we are not sure whether the ordering of the indices matter, we should not write, for instance τ_b^{ac} because this could mean, say $\tau_c^{a \ b}$ or τ_b^{ac} .

6 Gravitation and Curvature

As we have said many times before, in general relativity, gravity is not a force. Rather, matter (mass and energy) causes spacetime to curve. And because gravity is not a force, we define **free particles** - objects who are not acted upon by any force - as objects moving with no forces **other than gravity**. The trajectory of free particles are called **geodesics**.

In this section, we will learn about:

- Curvature: How can we tell if a space is curved?
- Geodesics: What is an equation to solve for a geodesic, and how can we solve for a geodesic?
- Motion in curved spaces: How do vectors behave as we move them along a geodesic? What if we move them not along a geodesic?
- The laws of physics: How do we describe the laws of physics in tensor form? As space can be curved and vectors can change due to the curvature of space, how we define the “derivative?”
- Newtonian limit: Will we be able to get back Newtonian physics in the limit?

This section will provide us with some powerful mathematical machinery for working with and understanding the Einstein equations and cosmology in the future.

6.1 Curvature

As tiny humans living of the surface of a giant ball, how could we tell that the space we are living in, where it seems locally flat, is actually curved? And what we mean by “going straight” once we have found out the Earth is round?

One way to test if a 2-dimensional space (the Earth for instance) is to just to walk, but with equal left and right step lengths. Have two people start walking parallel to each other. If one point their paths cross, then the space is curved.

Another way to see if the Earth is curved or flat is doing the “triangle test.” Simply generated a “spherical triangle” on the surface of Earth by intersecting three **great circles** (so the equator and two perpendicular (at the pole) longitude lines would suffice). We would observe that the sum of the angles in this triangle is 270° , not 180° . This says that the surface of the Earth is curved.

Now, before we go on and attempt to make a rigorous procedure to distinguish curved and flat spaces, we should make a few key observations this above example:

- If parallel lines cross in a space, then the space is non-Euclidean and is therefore curved.
- Lines formed when walking “straight” (walking with left step size equal right step size as in our first method) are geodesics.
- On a sphere, great circles are circles whose center is the origin, for example: the equator, longitude lines. Beware that latitude lines are not great circles. On a sphere, great circles are geodesics.
-

6.2 Geodesics and Affine connections $\Gamma_{\mu\nu}^{\sigma}$

In this section, we will learn a procedure to generate a geodesic, given that we are in a space or spacetime where we know the metric of the space. This procedure will follow something along the line of “moving straight” in the space, but much more mathematically rigorous.

6.2.1 Flat 3D space

Let us consider the simplest possible case with flat 3D case where geodesics are, as we guessed it, literally straight lines. In Cartesian coordinates, a straight line obeys the “no-curvature” equation:

$$\frac{\partial^2 \vec{r}}{\partial t^2} = \vec{0}.$$

Now, suppose that we are in an arbitrary curvilinear coordinate system. What would be the equation of a straight line here? It makes sense to think about **distance** and **displacement**. If we are moving such that the distance covered the same as displacement, at every instance, then we are moving in a straight line. So, let our displacement be \vec{r} . With arc length parameterization: $\vec{r} = \vec{r}(s)$, where s is essentially distance traveled. The condition of straightness gives:

$$\left\| \frac{d\vec{r}}{ds} \right\| = 1$$

We should recognize that $d\vec{r}/ds$ is nothing but the tangent vector to the path. Let's call this vector $\vec{\lambda}$. Using the chain rule, it follows that

$$\vec{\lambda} = \frac{d\vec{r}}{ds} = \frac{\partial \vec{r}}{\partial u^i} \frac{\partial u^i}{\partial s} = \frac{\partial u^i}{\partial s} \vec{e}_i = \lambda^i \vec{e}_i,$$

where u^i is simply a coordinate in the curvilinear coordinate system. The above equality gives:

$$\lambda^i = \frac{\partial u^i}{\partial s} = \dot{u}^i(s)$$

Now, be $\vec{\lambda}$ is a tangent vector, its direction does not change along a straight line. Since we also know that $||\vec{r}(s)|| = 1$, $\vec{\lambda}$ has both fixed direction and magnitude along a straight line. Therefore, “straightness,” at least in this limited case, implies that the derivative of tangent vectors with respect to arc length is zero. We call this implication the **condition of straightness**:

$$\boxed{\frac{\partial \vec{\lambda}}{\partial s} = 0}$$

Let’s take a closer look at the boxed expression but in a different coordinate system. If we think about the above equality, it should also make sense for any general 3-dimensional coordinate system in flat space, in which case:

$$\frac{d}{ds} (\lambda^i \vec{e}_i) = \dot{\lambda}^i \vec{e}_i + \lambda^i \dot{\vec{e}}_i = \vec{0}.$$

In Cartesian coordinates, the basis vectors \vec{e}_i happen to be constant, so $\dot{\vec{e}}_i = \vec{0}$. This gives $\dot{\lambda}^i = 0$. And since $\lambda^i = \dot{u}^i = \dot{x}^i$, which implies $\partial^2 x^i / \partial s^2 = \vec{0}$. Now imagine if we parameterize \vec{r} with a different parameter, say t . If $t \propto s$, then

$$\frac{\partial^2 x^i}{\partial s^2} = \frac{\partial^2 x^i}{\partial t^2} = \vec{0}$$

holds. This is equivalent to the case we have an object moving with constant velocity (so that $s = vt$). If s is not proportional to t then the above equality does not hold.

We have established the “condition of straightness” in Cartesian coordinates. However, the equation $\partial^2 x^i / \partial s^2 = 0$ does not hold for general coordinate systems. Rather, the more fundamental equation applies:

$$\boxed{\frac{\partial \lambda^i}{\partial s} \vec{e}_i + \lambda^i \frac{\partial \vec{e}_i}{\partial s} = 0}$$

where

$$\frac{\partial \vec{e}_i}{\partial s} = \frac{\partial \vec{e}_i}{\partial u^j} \frac{\partial u^j}{\partial s} = \frac{\partial \vec{e}_i}{\partial u^j} \dot{u}^j \neq 0$$

in general. Now, to simplify our notation, let us use (as we have introduced before) $\partial_j = \partial / \partial u^j$. This gives:

$$\frac{\partial \lambda^i}{\partial s} \vec{e}_i + \lambda^i \frac{\partial \vec{e}_i}{\partial s} = \dot{\lambda}^i \vec{e}_i + \lambda^i (\partial_j \vec{e}_i) \dot{u}^j = 0$$

Note that $\partial_j \vec{e}_i$ are vectors, which we can expand in terms of basis vectors, but we will not do that. Rather, let us define the **Christoffel symbol**, or the **affine connection**, Γ_{ij}^k in terms of $\partial_j \vec{e}_i$ as:

$$\boxed{\partial_j \vec{e}_i = \Gamma_{ij}^k \vec{e}_k}$$

We will interpret Γ_{ij}^k as “the k^{th} component of the u^j derivative of \vec{e}_i .” It is very important that we keep in mind Γ_{ij}^k is **not** a tensor. Rather, the Christoffel symbols are **connections** - they are just vector components.

With the definition of Γ_{ij}^k , we can express the derivative of \vec{e}_i as:

$$\dot{\vec{e}}_i = (\partial_j \vec{e}_i) \dot{u}^j = \Gamma_{ij}^k \vec{e}_k \dot{u}^j.$$

It follows that the condition of straightness becomes:

$$\begin{aligned} \frac{d\vec{\lambda}}{ds} &= \dot{\lambda}^i \vec{e}_i + \lambda^i \Gamma_{ij}^k \vec{e}_k \dot{u}^j \\ &= \dot{\lambda}^i \vec{e}_i + \lambda^j \Gamma_{jk}^i \vec{e}_i \dot{u}^k, \text{ letting } i \rightarrow j, j \rightarrow k, k \rightarrow i \\ &= \left(\dot{\lambda}^i + \Gamma_{jk}^i \dot{u}^k \right) \vec{e}_i \\ &= 0, \end{aligned}$$

or

$$\dot{\lambda}^i + \Gamma_{jk}^i \dot{u}^k = 0.$$

Let us write the above equality more completely with the arc length parameter s , and call it the equation of a straight line in flat 3-dimensional space. But notice how the number of dimensions do not really play a role anywhere in our derivation. This hints to us that the equation also holds for n -dimensional manifolds, but we will discuss this topic later on.

$$\boxed{\frac{d^2 u^i}{ds^2} + \Gamma_{jk}^i \frac{du^j}{ds} \frac{du^k}{ds} = 0}$$

For i, j, k , the above expressions encodes 3 equations whose solution is geodesic in flat space.

Now, let us “worry” a little bit about the Christoffel symbol Γ_{jk}^i . It seems like that because Γ_{jk}^i has three indices i, j, k , there will be 27 coefficients associated with different combinations of i, j, k . However, we can show that this is not the case. Recall that

$$\partial_j \vec{e}_i = \Gamma_{ij}^k \vec{e}_k.$$

Dotting the entire equation with \vec{e}^l gives

$$\partial_j \vec{e}_i \cdot \vec{e}^l = \Gamma_{ij}^k \vec{e}_k \cdot \vec{e}^l = \Gamma_{ij}^k \delta_k^l = \Gamma_{ij}^l.$$

Also, observe that

$$\partial_j \vec{e}_i = \frac{\partial}{\partial u^j} \frac{\partial \vec{r}}{\partial u^i} = \frac{\partial}{\partial u^i} \frac{\partial \vec{r}}{\partial u^j} = \partial_i \vec{e}_j.$$

This simply implies a symmetric relation:

$$\boxed{\Gamma_{ij}^l = \Gamma_{ji}^l}$$

So, it turns out that we only have 18 independent terms, instead of 27.

Next, we want to the relationship between the Christoffel symbols and the metric tensor. Specifically, we want to find out how to compute the Christoffel symbols, using the metric tensor. Consider:

$$\begin{aligned}\partial_k g_{ij} &= \partial_k (\vec{e}_i \cdot \vec{e}_j) \\ &= \vec{e}_j \partial_k \vec{e}_i + \vec{e}_i \partial_k \vec{e}_j \\ &= \vec{e}_j \Gamma_{ik}^m \vec{e}_m + \vec{e}_i \Gamma_{kj}^m \vec{e}_m.\end{aligned}$$

This gives

$$\boxed{\partial_k g_{ij} = \Gamma_{ik}^m g_{jm} + \Gamma_{jk}^m g_{im}}$$

By symmetry, we also get

$$\boxed{\partial_i g_{jk} = \Gamma_{ji}^m g_{km} + \Gamma_{ki}^m g_{jm}}$$

and

$$\boxed{\partial_j g_{ik} = \Gamma_{kj}^m g_{im} + \Gamma_{ij}^m g_{km}}$$

Adding the first two equations and subtract the third, we get

$$\partial_k g_{ij} + \partial_i g_{jk} - \partial_j g_{ik} = 2\Gamma_{ik}^m g_{jm}.$$

Next, multiplying this entire equation by g_{jl} , so that $g_{jm}g^{jl} = \delta_m^l$, we get

$$\Gamma_{ik}^l = \frac{1}{2} g^{jl} (\partial_k g_{ij} + \partial_i g_{jk} - \partial_j g_{ik}).$$

Finally, letting $l \rightarrow k, k \rightarrow i, i \rightarrow j, j \rightarrow l$, we obtain the formula to compute the Christoffel symbol from the metric:

$$\boxed{\Gamma_{ij}^k = \frac{1}{2} g^{kl} (\partial_i g_{jl} + \partial_j g_{il} - \partial_l g_{ij})}$$

Now, in Cartesian coordinates, the metric is simply the identity matrix, which is constant. This simply means that all Christoffel symbols Γ_{ij}^k in Cartesian coordinates are zero. However, this does not imply if the Christoffel symbols are non-zero then the space is curved. In fact, we get non-zero Christoffel symbols in curvilinear coordinates in flat space whenever the basis vectors \vec{e}_i are not constant. An example would be spherical coordinates, which we will soon find

out.

We might ask whether there is a nice, short way to calculate the Christoffel symbols. Unfortunately, the answer is no - brute force is the only way to go. Luckily, though, Mathematica can handle Christoffel symbol computations, given a metric tensor.

Example 6.1. Find $\Gamma_{23}^1 = \Gamma_{32}^1$.

Again, brute force is the way to go:

$$\begin{aligned}\Gamma_{23}^1 &= \Gamma^{132} \\ &= \frac{1}{2}g^{11}(\partial_2 g_{31} + \partial_3 g_{21} - \partial_1 g_{23}) \\ &\quad + \frac{1}{2}g^{12}(\partial_2 g_{32} + \partial_3 g_{22} - \partial_2 g_{23}) \\ &\quad + \frac{1}{2}g^{13}(\partial_2 g_{33} + \partial_3 g_{23} - \partial_3 g_{23})\end{aligned}$$

We can of course repeat this process to find the other 25 terms, but we will not do that!

We have used arc length as a parameter in finding the geodesic equation:

$$\frac{d^2 u^i}{ds^2} + \Gamma_{jk}^i \frac{du^j}{ds} \frac{du^k}{ds} = 0.$$

Now, what if we use a different parameter $t = f(s)$? How will the form of the equation above change? We can show (but we won't) that with $t = f(s)$, the above equation has an extra term:

$$\boxed{\frac{d^2 u^i}{ds^2} + \Gamma_{jk}^i \frac{du^j}{ds} \frac{du^k}{ds} = -\frac{d^2 t}{ds^2} \left(\frac{dt}{ds}\right)^{-2} \frac{du^i}{dt}}$$

But if we have a linear relationship between t and s , i.e.,

$$t = As + B,$$

where $A \neq 0$ and B are constants, then we get

$$\boxed{\frac{d^2 u^i}{dt^2} + \Gamma_{jk}^i \frac{du^j}{dt} \frac{du^k}{dt} = -\frac{d^2 t}{ds^2} \left(\frac{dt}{ds}\right)^{-2} \frac{du^i}{dt} = 0}$$

simply because

$$\frac{d^2 t}{ds^2} = 0.$$

A parameter t of this form $t = As + B$ is called an **affine parameter**.

6.3 Geodesics in curved space

We have seen multiple correspondences between flat 3-dimensional space in curvilinear coordinates and curved general N -dimensional manifolds:

$$\begin{aligned} u^j &\leftrightarrow x^a \\ g_{ij} &\leftrightarrow g_{ab} \\ \lambda^{i'} &= U_j^{i'} \lambda^j \leftrightarrow \lambda^{a'} X_b^{a'} \lambda^b \\ ds^2 &= g_{ij} du^i du^j \leftrightarrow ds^2 = g_{ab} du^a du^b, \end{aligned}$$

which is not surprising because flat 3-dimensional space is simply a special case of general curved N -dimensional spaces. We have expected the form of the geodesic equation in flat 3-dimensional space to remain the same in general curved N -dimensional space. And we should be correct, because nowhere in our derivation of the geodesic equation is dimension-specific, i.e., the derived equation does not just apply exclusively to the flat three dimensional case. So, the geodesic equation in general curved N -dimensional space is simply

$$\boxed{\frac{d^2 x^a}{d\sigma^2} + \Gamma_{bc}^a \frac{dx^b}{d\sigma} \frac{dx^c}{d\sigma} = 0}$$

where σ is an **affine parameter**. The Christoffel symbol Γ_{bc}^a is defined in exactly the same fashion as before:

$$\boxed{\Gamma_{bc}^a = \frac{1}{2} g^{ad} (\partial_b g_{cd} + \partial_c g_{bd} - \partial_d g_{bc})}$$

Now, let us dwell into a nice example that illustrates how the geodesic equation gives a geodesic.

Example 6.2. Determine if lines of constant latitude of a 2-sphere of radius a are geodesics.

We have said before that only great circles on a sphere are geodesics. Now, we can verify this fact using the geodesic equation. First, consider any circle on the 2-sphere of radius 2. Let us oriented the sphere so that this line is the line of constant latitude.

insert figure 16 here!!!

Next, let us gather all we have known about the geometry of a 2-sphere. The metric tensor is given by

$$[g_{AB}] = \begin{pmatrix} a^2 & 0 \\ 0 & a^2 \sin^2 \theta \end{pmatrix},$$

and the inverse metric tensor is given by

$$[g^{AB}] = \begin{pmatrix} a^{-2} & 0 \\ 0 & a^{-2} \sin^{-2} \theta \end{pmatrix},$$

where θ is the angle form by the position vector \vec{r} with the z -axis. The line of constant latitude is given by the parameters:

$$u^A = (u^1, u^2) = (\theta, \phi),$$

where θ is fixed and $0 \leq \phi \leq 2\pi$. The question we should be asking ourselves is whether

$$\frac{d^2 u^A}{ds^2} + \Gamma_{BC}^A \frac{du^B}{ds} \frac{du^C}{ds} = 0,$$

where s is the arc length parameter, holds for certain values of θ . Because lines of constant latitudes are circles, the arc length s is easy to find: $s(\phi) = r'\phi = a\phi \sin \theta$. This gives:

$$\phi = \frac{s}{a \sin \theta}.$$

Therefore, we can parameterize the path with the arc length s :

$$u^A(s) = (u^1, u^2)(s) = \left(\theta, \frac{s}{a \sin \theta} \right).$$

But since θ is fixed,

$$\frac{d^2 u^A}{ds^2} = (0, 0).$$

So it seems that we only need to check for which θ

$$\Gamma_{BC}^A \frac{du^B}{ds} \frac{du^C}{ds} = 0.$$

So the next reasonable step to take is computing the Christoffel symbols. There are 8 of these, but it turns out (by symmetry and the fact that g_{AB} is diagonal), we only have to do a few computations:

$$\begin{aligned} \Gamma_{22}^1 &= -\sin \theta \cos \theta \\ \Gamma_{12}^2 &= \Gamma_{21}^2 = \cot \theta \\ \Gamma_{12}^1 &= \Gamma_{21}^1 = \Gamma_{11}^2 = \Gamma_{22}^2 = 0. \end{aligned}$$

For $A = 1$,

$$\begin{aligned} \Gamma_{BC}^A \frac{du^B}{ds} \frac{du^C}{ds} &= 0 \\ \Gamma_{BC}^1 \frac{du^B}{ds} \frac{du^C}{ds} &= 0 \\ -\sin \theta \cos \theta \frac{du^2}{ds} \frac{du^2}{ds} &= 0 \\ -\sin \theta \cos \theta (a \sin \theta)^{-2} &= 0 \end{aligned}$$

can only be true if $\theta = \pi/2$. This suggests that the equator is the only viable candidate to be a geodesic. However, we should check if the equality can hold for $A = 2$, because if the equality cannot hold then we have to conclude that lines of constant latitude cannot be geodesics. Let us check:

$$\begin{aligned}\Gamma_{BC}^A \frac{du^B}{ds} \frac{du^C}{ds} &= 0 \\ \Gamma_{BC}^2 \frac{du^B}{ds} \frac{du^C}{ds} &= 0 \\ (0 + 0) \cot \theta &= 0.\end{aligned}$$

Ah! So equality holds for any θ in this case. Combining the above two facts, we conclude that only the circle of latitude $\pi/2$ works as a geodesic. Equivalently, for a sphere, only great circles are geodesics.

6.4 Parallel transport

Our condition for constructing a geodesic was that the tangent vector $\vec{\lambda} = \lambda^i \vec{e}_i = \dot{\lambda}_i = \dot{e}^i = \dot{u}^i \vec{e}_i$ does not change as we move along the curve, i.e. the condition of straightness:

$$\frac{d\vec{\lambda}}{ds} = 0,$$

which led us to the geodesic equation:

$$\ddot{u}^i + \Gamma_{jk}^i \dot{u}^j \dot{u}^k = 0.$$

Now, what about any arbitrary vector $\vec{\lambda}$ that is not a tangent to the geodesic? To ask how $\vec{\lambda}$ changes as it moves along a geodesic is a legitimate question. However, we have to be more specific. There are infinitely many ways for $\vec{\lambda}$ to travel along the geodesic because the “tangent condition” is no longer required. So, let us require that we move $\vec{\lambda}$ along the geodesic in such a way that the only changes happening to the $\vec{\lambda}$ are due to the geometry of the space. For instance, imagine pushing a pencil along the equator without rotating it. Obviously the direction of the pencil changes over time, but not because we are exerting a torque on it but rather because the surface of the globe is a curved space. We this action of “moving a vector in a space without changing it” **parallel transport**. We will define this term more rigorously as we move on.

Now, because $\vec{\lambda}$ does not change. Let t be an affine parameter, we get the **condition of parallel transport**:

$$\frac{d\vec{\lambda}}{dt} = 0.$$

In flat space, $\vec{\lambda}$ does not change its direction due to parallel transport. But in curved spaces, as we have discussed in the example of the pencil and the globe,

$\vec{\lambda}$ can change directions due to curvature. Now, we can expand the condition of parallel transport to get further insights:

$$\begin{aligned}\frac{d\vec{\lambda}}{dt} &= 0 \\ \dot{\lambda}^i \vec{e}_i + \lambda^i \dot{\vec{e}}_i &= 0.\end{aligned}$$

Since we also know that

$$\dot{\vec{e}}_i = \dot{u}^j \partial_j \vec{e}_i = \Gamma_{ij}^k \dot{u}^j \vec{e}_k,$$

letting $k \rightarrow i, i \rightarrow j, j \rightarrow k$, we get

$$\dot{\lambda}^i + \lambda^j \Gamma_{kj}^i \dot{u}^k = 0$$

This equation tells us how the components λ^i change when $\vec{\lambda}$ is parallel transported along any curve parameterized by the affine parameter t . Note (again) that our geodesic equations are not dimension-specific, i.e., we can just change the indices for the equation to work in any N -dimensional manifold:

$$\boxed{\dot{\lambda}^a + \lambda^b \Gamma_{bc}^a \dot{u}^c = 0}$$

Example 6.3. Show that we get the geodesic equation from the above equality in the case where $\vec{\lambda}$ is a tangent vector.

If $\lambda^i = \dot{u}^i$, then the above equality simply becomes:

$$\ddot{u}^i + \Gamma_{jk}^i \dot{u}^j \dot{u}^k = 0,$$

which is exactly the geodesic equation. Note that this also says to parallel transport tangent vectors, the curve must be a geodesic (so that $\vec{\lambda}$ remains a tangent vector).

Example 6.4. Consider a unit vector $\vec{\lambda}$ on the surface of a sphere of radius a which makes an angle α with respect to a longitude line.

insert figure 17 here!!!

Show that as $\vec{\lambda}$ is parallel transported along the line of constant latitude θ_0 , the direction of $\vec{\lambda}$ changes by an angle $\chi = 2\pi\omega$, where $\omega = \cos \theta_0$.

First, we parameterize the curve as $u^A = (u^1, u^2) = (\theta_0, \phi)$, where θ_0 is fixed and $0 \leq \phi \leq 2\pi$. The arc length is $s = \phi a \sin \theta_0$. So this gives, as before, $\phi = s (a \sin \theta_0)^{-1}$. Next, let $\vec{\lambda}(0)$ be the vector before parallel transport and $\vec{\lambda}(2\pi)$ be the vector after parallel transport. We know that χ is the angle between formed between these two vectors. Now, because $\vec{\lambda}(0)$ makes an angle α with a longitude line, we have

$$\lambda^A(0) = (\lambda^1(0), \lambda^2(0)) = \left(a^{-1} \cos \alpha, (a \sin \theta_0)^{-1} \sin \alpha \right).$$

We can readily verify that $\vec{\lambda}(0)$ is indeed a unit vector:

$$\begin{aligned} & g_{AB}\lambda^A(0)\lambda^B(0) \\ &= \begin{pmatrix} a^{-1}\cos\alpha & (a\sin\theta_0)^{-1}\sin\alpha \end{pmatrix} \begin{pmatrix} a^2 & 0 \\ 0 & (a\sin\theta_0)^2 \end{pmatrix} \begin{pmatrix} a^{-1}\cos\alpha \\ (a\sin\theta_0)^{-1}\sin\alpha \end{pmatrix} \\ &= \cos^2\alpha + \sin^2\alpha = 1. \end{aligned}$$

We can also readily verify that $\lambda^A(0)$ makes an angle α with the longitude line. Let μ^A be a vector that points along a longitude line. In Cartesian coordinates,

$$\mu^A = 1\vec{e}_\theta + 0\vec{e}_\phi.$$

So, in the spherical basis, μ^A simply becomes $(1, 0)^\top$. The angle between λ^A and μ^A can come out of their inner product:

$$\cos\beta = \frac{g_{AB}\mu^A\lambda^B}{\|\mu^A\|\|\lambda^B\|} = \frac{a^2a^{-1}\cos\alpha}{a} = \cos\alpha.$$

So, the angle between μ^A and λ^A is indeed α . Next, we parallel transport λ^A along the line of constant latitude θ_0 . We need to solve the parallel transport equations (there will be two of them) to find the new components of $\vec{\lambda}$ after parallel transporting:

$$\dot{\lambda}^A + \Gamma_{BC}^A\lambda^B\dot{u}^C = 0.$$

With initial values

$$\vec{\lambda}(0) = \begin{pmatrix} a^{-1}\cos\alpha \\ (a\sin\theta_0)^{-1}\sin\alpha \end{pmatrix},$$

the problem boils down to an ordinary differential initial value problem. Using the Christoffel symbols we have found before, we have a coupled system of equations:

$$\begin{cases} \dot{\lambda}^1 + \Gamma_{22}^1\lambda^2\dot{u}^2 = 0 \\ \dot{\lambda}^2 + \Gamma_{12}^2\lambda^1\dot{u}^1 + \Gamma_{21}^2\lambda^2\dot{u}^2 = 0 \end{cases}.$$

We are not going over the details of how to solve this system, but we certainly verify that

$$\begin{aligned} \lambda^A(t) &= (\lambda^1(t), \lambda^2(t)) \\ &= \left(a^{-1}\cos(\alpha - \omega t), (a\sin\theta_0)^{-1}\sin(\alpha - \omega t) \right) \end{aligned}$$

where $\omega = \cos\theta_0$ solves the initial value problem. Now, let $t = 2\pi$, we get

$$\lambda^A(2\pi) = \left(a^{-1}\cos(\alpha - 2\pi\omega), (a\sin\theta_0)^{-1}\sin(\alpha - 2\pi\omega) \right).$$

The final step is to find the angle between $\lambda^A(0)$ and $\lambda^A(2\pi)$:

$$\begin{aligned}\cos \chi &= \frac{g_{AB}\lambda^A(0)\lambda^B(2\pi)}{||\lambda^A(0)||||\lambda^B(2\pi)||} = \frac{g_{AB}\lambda^A(0)\lambda^B(2\pi)}{1} \\ &= a^2 a^{-1} \cos \alpha a^{-1} \cos(\alpha - 2\pi\omega) + (a \sin \theta_0)^2 (a \sin \theta_0)^{-2} \sin \alpha \sin(\alpha - 2\pi\omega) \\ &= \cos \alpha \cos(\alpha - 2\pi\omega) + \sin \alpha \sin(\alpha - 2\pi\omega) \\ &= \cos(\alpha - \alpha + 2\pi\omega) \\ &= \cos 2\pi\omega\end{aligned}$$

So, $\chi = 2\pi\omega = 2\pi \cos \theta_0$. What happens if we look at $\theta_0 = \pi/2$? Not surprisingly, $\chi = 2\pi \cos \omega = 2\pi \cos 2\pi = 0$, i.e., the direction of $\lambda^A(2\pi)$ is the same as $\lambda^A(0)$ if we move it along a geodesic.

6.5 Curved Spacetime

As we have established before, the geodesic equations are not dimension-specific. So we expect that the form of the geodesic equations hold in curved spacetime, and we would be right:

$$\frac{d^2 x^\mu}{d\tau^2} + \Gamma_{\nu\sigma}^\mu \frac{dx^\nu}{d\tau} \frac{dx^\sigma}{d\tau} = 0.$$

with the Christoffel symbols defined in the same fashion as before

$$\Gamma_{\nu\sigma}^\mu = \frac{1}{2} g^{\mu\rho} (\partial_\nu g_{\sigma\rho} + \partial_\sigma g_{\nu\rho} - \partial_\rho g_{\nu\sigma}).$$

Instead of giving us the trajectory of a free particle in space, the geodesic equations in curved spacetime gives the trajectories of free particles in spacetime $x^\mu(\tau)$. For example, we can solve for the trajectory of a particle in a gravitational field (recall that in general relativity, gravity is not a force). Also recall that for massive particles, we can use the proper time τ as an affine parameter due to the relation $ds^2 = c^2 d\tau^2$.

Likewise, any vector λ^μ can be parallel transported along a curve $x^\mu(\tau)$ where how the components changes obeys the 4-dimensional parallel transport equations:

$$\dot{\lambda}^\mu + \Gamma_{\nu\sigma}^\mu \lambda^\nu \dot{x}^\sigma = 0.$$

6.6 Principle of covariance

The next question we want to ask ourselves is, given the geometry of spacetime and the trajectory of free particles, can we formulate the laws of physics for curved spacetime in a way that is the law of physics in flat spacetime and flat space are just special cases. Of course, to answer this question we require the **principle of covariance**.

Recall that one of the postulates of special relativity is that the laws of physics are the same in all inertial frames, i.e., equations of physics are **invariant** under general Lorentz transformations.

Example 6.5.

$$\begin{aligned} f^\mu &= \frac{dp^\mu}{d\tau} \\ \Lambda_{\mu}^{\nu'} f^\mu &= \Lambda_{\mu}^{\nu'} \frac{dp^\mu}{d\tau} \\ f^{\nu'} &= \frac{d}{d\tau} \left(\Lambda_{\mu}^{\nu'} p^\mu \right) \\ f^{\nu'} &= \frac{dp^{\nu'}}{d\tau}. \end{aligned}$$

The first equality is equivalent to the last equality. In fact, they differ only in the dummy indices. Note that the metric tensors are also the same in both frames $\eta_{\mu\nu} = \eta_{\mu'\nu'}$, we can indeed confirm that the two mentioned equalities are the same. This example is one many verifications for the fact that equations of physics in special relativity are invariant under Lorentz transformations.

In general relativity, equations have to maintain the same form under general coordinate transformations. We have been aware that the metric tensor and Christoffel symbols can differ in different coordinate systems, so we are not guaranteed the invariant condition. Therefore, in general relativity, while the laws of physics do not have to be **invariant** under transformations, they have to be **covariant**. Note that invariance implies covariance. But the converse is not true.

In trying to figure out how equations hold in curved spacetime, Einstein introduced the **principle of covariance**:

An equation is true in general relativity in all coordinate systems if:

- It is true in special relativity.
- It is a **tensor equation** that preserves its form under general coordinate transformations (covariance).

We should make a few remarks on this principle. The first condition stems from the equivalence principle, which states that there is always a freely falling frame where the laws of special relativity hold locally. The second condition is motivated by the fact that tensors of the same type all transform the same way under general coordinate transformations. For example, for tensors $A^\mu = B^\mu$,

$$A^{\mu'} = X_n^{\mu'} u A^\nu = X_\nu^{\mu'} B^\nu = B^{\mu'}.$$

What the principle of covariance is saying is that as long as the laws of special relativity are written in tensor form, the same equations will be true in the presence of gravity. This gives us a powerful prescription for finding the laws of physics in general relativity. Let's us solidify our understanding so far by looking an example where things go wrong.

Example 6.6. We know that

$$f^\mu = \frac{dp^\mu}{d\tau}$$

holds in special relativity (flat spacetime). Does this equation also hold in general curved spacetime?

We shall check whether the equation above is a tensor equation. If it is, then yes, if not then no. We should not have any difficulty in recognizing that $dp^\mu/d\tau$ is not a tensor because it does not transform correctly under general coordinate transformations. We can show this simply by applying the chain rule:

$$\frac{dp^\mu}{d\tau} = \frac{d}{d\tau} (X_{\nu'}^{\mu'} p^{\nu'}) = X_{\nu'}^{\mu'} \frac{dp^{\nu'}}{d\tau} + p^{\nu'} \frac{dX_{\nu'}^{\mu'}}{d\tau} \neq X_{\nu'}^{\mu'} \frac{dp^{\nu'}}{d\tau}$$

in general because the Jacobian matrix element $X_{\nu'}^{\mu'}$ is not necessarily independent of the proper time τ . Therefore,

$$f^\mu = \frac{dp^\mu}{d\tau}$$

is not covariant, and therefore it does not apply to a general coordinate system.

Example 6.7. Given the metric tensor $g_{\mu\nu}$. Show that $\partial_\lambda g_{\mu\nu}$ is not a tensor.

We simply apply the transformation rule

$$\partial_{\lambda'} g_{\mu'\nu'} = \partial_{\lambda'} X_{\mu'}^\alpha X_{\nu'}^\beta g_{\alpha\beta} \neq X_{\lambda'}^\gamma X_{\mu'}^\alpha X_{\nu'}^\beta \partial_\gamma g_{\alpha\beta}.$$

So, $\partial_\lambda g_{\alpha\beta}$ is not a tensor. For this reason,

$$\Gamma_{\lambda\nu}^\mu = \frac{1}{2} g^{\mu\rho} (\partial_\nu g_{\rho\lambda} + \partial_\lambda g_{\gamma\rho} - \partial_\rho g_{\nu\lambda})$$

is also not a tensor, even though it is covariant, i.e. it retains the same form in a primed frame.

$$\Gamma_{\lambda'\nu'}^{\mu'} = \frac{1}{2} g^{\mu'\rho'} (\partial_{\nu'} g_{\rho'\lambda'} + \partial_{\lambda'} g_{\gamma'\rho'} - \partial_{\rho'} g_{\nu'\lambda'})$$

The above examples suggest to us that “things go wrong” because of the derivative ∂_τ . In fact, we have just shown that the derivative of a tensor is not

necessarily a tensor under general coordinate transformations. So, it is reasonable to suspect that in order to derive a procedure for finding the laws of physics in a different coordinate system, we would first need to generalize our current definition of the derivative so that derivatives of tensors are tensors.

Let us revisit the our current definition of the derivative and see where we would run into trouble. Consider a vector λ^a in a general curved space. Let λ^a at P and Q denote how λ^a evolves over some Δt .

insert figure here!!!

$$\frac{d\lambda^a}{dt} = \lim_{\Delta t \rightarrow 0} \frac{\lambda^a(t + \Delta t) - \lambda^a(t)}{\Delta t}.$$

Applying the transformation rules $X_a^{b'}$ on λ^a at point Q and $X_a^{b'}(t + \Delta t)$ on λ^a at P , we realize that the Jacobian matrix element $X_a^{b'}$ can depend on t , i.e., the space is different at P and Q , so it does make sense to compare λ^a at P to λ^a at Q if we cannot account for the change in the space. A better approach would be to compare $\lambda^a(t + \Delta t)$ and λ^a at the same point. And to do this, we need to parallel transport $\lambda^a(t + \Delta t)$ from Q to P . We shall explore the ‘new’ definition of the derivative in the next section.

6.7 Absolute and Covariant differentiation

6.7.1 Absolute differentiation

Consider a manifold and a contravariant vector λ^a parameterized by t . It follows from the definition of the derivative with respect to t that:

$$\frac{d\lambda^a}{dt} = \lim_{\Delta t \rightarrow 0} \frac{\lambda^a(\Delta t + t) - \lambda^a(t)}{\Delta t}.$$

We have stated that we run into trouble with this definition because the space at P and Q are somehow “different” due to general curvature, i.e.

$$X_{b'}^{a'}(P) \neq X_b^{a'}(Q).$$

Specifically, as $\Delta t \rightarrow 0$, we will also get extra terms that are derivatives of $X_{b'}^{a'}$. To ‘fix’ this issue, we introduce the **absolute derivative**:

$$\frac{D\lambda^a}{dt} = \lim_{\Delta t \rightarrow 0} \frac{\lambda^a(\Delta t + t) - \bar{\lambda}^a}{\Delta t},$$

where $\bar{\lambda}^a = \lambda^a$ at point P but parallel transported to point Q . Now, we want an explicit expression for the absolute derivative. To do this, we Taylor expand the first term in the definition

$$\lambda^a(\Delta t + t) \approx \lambda^a(t) + \frac{d\lambda^a}{dt} \Delta t = \lambda^a(P) + \frac{d\lambda^a}{dt} \Delta t,$$

and apply the parallel transport equation to the second term:

$$\dot{\lambda}^a + \Gamma_{bc}^a \lambda^b \dot{x}^c = 0.$$

For small finite intervals,

$$\begin{aligned}\dot{\lambda}^a &\approx \frac{\Delta \lambda^a}{\Delta t} \\ \dot{x}^c &\approx \frac{\Delta x^c}{\Delta t}.\end{aligned}$$

So the parallel transport equation becomes:

$$\Delta \lambda^a + \Gamma_{bc}^a \lambda^b \Delta x^c = 0,$$

where $\Delta \lambda^a = \bar{\lambda}^a(Q) - \lambda^a(P)$. This gives

$$\bar{\lambda}^a(Q) \approx \lambda^a(P) - \Gamma_{bc}^a \lambda^b \Delta x^c.$$

Plugging this into the definition of the absolute derivative, we get

$$\begin{aligned}\frac{D\lambda^a}{dt} &= \lim_{\Delta t \rightarrow 0} \frac{(d\lambda^a/dt)\Delta t + \Gamma_{bc}^a \lambda^b \Delta x^c}{\Delta t} \\ &= \lim_{\Delta t \rightarrow 0} \left(\frac{d\lambda^a}{dt} + \Gamma_{bc}^a \lambda^b \frac{\Delta x^c}{\Delta t} \right) \\ &= \frac{d\lambda^a}{dt} + \Gamma_{bc}^a \lambda^b \dot{x}^c\end{aligned}$$

So, let us define the absolute derivative of a contravariant vector as:

$$\boxed{\frac{D\lambda^a}{dt} = \frac{d\lambda^a}{dt} + \Gamma_{bc}^a \lambda^b \dot{x}^c}$$

Note that this definition is covariant as expected:

$$\frac{D\lambda^a}{dt} = X_{b'}^a \frac{D\lambda^{b'}}{dt}.$$

We also notice that the right hand side of the definition is quite similar to the parallel transport equation. In fact,

$$\frac{d\lambda^a}{dt} + \Gamma_{bc}^a \lambda^b \dot{x}^c = 0$$

if we parallel transport a contravariant vector λ^a . This says that λ^a is constant under absolute differentiation. We get a nice property for a parallel transported λ^a :

$$\boxed{\frac{D\lambda^a}{dt} = 0}$$

A reasonable question to ask now is what the absolute differentiation of scalars, covariant vectors, and general tensors look like.

Recall that scalars are invariant under general coordinate transformations, and that invariance implies covariance. This implies absolute derivatives of scalars are just the normal derivatives of scalars:

$$\boxed{\frac{D\phi}{dt} = \frac{d\phi}{dt}}$$

To find the form of the absolute derivatives of covariant vectors, we first consider the scalar $\lambda^a \mu_a$. By the definition of absolute derivatives of scalars, we get:

$$\begin{aligned} \frac{D\lambda^a \mu_a}{dt} &= \frac{\lambda^a \mu_a}{dt} \\ \frac{D\lambda^a}{dt} \mu_a + \frac{D\mu_a}{dt} \lambda^a &= \frac{d\lambda^a}{dt} \mu_a + \frac{d\mu_a}{dt} \lambda^a \\ \mu_a \left(\frac{d\lambda^a}{dt} + \Gamma_{bc}^a \lambda^b \dot{x}^c \right) + \lambda^a \frac{D\mu_a}{dt} &= \frac{d\lambda^a}{dt} \mu_a + \lambda^a \frac{d\mu_a}{dt} \\ \frac{D\mu_a}{dt} &= \frac{1}{\lambda^a} \left(\lambda^a \frac{d\mu_a}{dt} - \mu_a \Gamma_{bc}^a \lambda^b \dot{x}^c \right) \\ \frac{D\mu_a}{dt} &= \frac{1}{\lambda^a} (\lambda^a - \mu_d \Gamma_{ac}^d \lambda^a \dot{x}^c). \end{aligned}$$

So,

$$\boxed{\frac{D\mu_a}{dt} = \frac{d\mu_a}{dt} - \Gamma_{ac}^d \mu_d \dot{x}^c} \quad (1)$$

We notice that the form of the absolute derivative of covariant vectors is similar to that of contravariant vectors except for the leading sign of the Christoffel symbol.

Lastly, combining what we know about absolute derivatives of contravariant and covariant objects, we can look at the absolute derivative of a general tensor. Consider $\tau^{ab}_c = \lambda^a \sigma^b \mu_c$ (recall that multiplying vectors components can give tensors). We can show that

$$\boxed{\frac{D\tau^{ab}_c}{dt} = \frac{d\tau^{ab}_c}{dt} + \Gamma_{de}^a \tau^{db}_c \dot{x}^e + \Gamma_{de}^b \tau^{ad}_c \dot{x}^e - \Gamma_{ce}^d \tau^{ab}_d \dot{x}^e}$$

Because absolute derivatives are tensors (by definition), we can apply the transformation rules as usual:

$$\frac{D\tau^{a'b'}_{c'}}{dt} = X_d^{a'} X_e^{b'} X_{c'}^f \frac{D\tau^{de}_f}{dt}$$

6.7.2 Covariant differentiation

The absolute derivative is taken with respect to a parameter, but we also need to take derivatives with respect to coordinates. In the latter case, we need to define the **covariant derivative** based on the “old” notion of the derivatives with respect to coordinates, which is

$$\partial_a = \frac{\partial}{\partial x^a}.$$

Let us define the covariant derivative by applying a chain rule to the absolute derivative, letting $\lambda^a = \lambda^a(t)$

$$\frac{D\lambda^a}{dt} = \frac{d\lambda^a}{dt} + \Gamma_{bc}^a \lambda^b \dot{x}^c = \frac{D\lambda^a}{dx^c} \frac{dx^c}{dt} = \frac{D\lambda^a}{dx^c} \dot{x}^c.$$

So,

$$\frac{D\lambda^a}{dx^c} \dot{x}^c = \frac{d\lambda^a}{dt} + \Gamma_{bc}^a \lambda^b \dot{x}^c = \frac{\partial \lambda^a}{\partial x^c} \frac{\partial x^c}{\partial t} + \Gamma_{bc}^a \lambda^b \dot{x}^c,$$

which gives the **covariant derivative** of λ^a with respect to x^c :

$$\boxed{\frac{D\lambda^a}{dx^c} = \frac{\partial \lambda^a}{\partial x^c} + \Gamma_{bc}^a \lambda^b = \partial_c \lambda^a + \Gamma_{bc}^a \lambda^b}$$

For notation simplicity sake, we define the “comma” and “semi-colon” derivatives for normal and covariant/absolute derivatives of covariant vectors. The semi-colon is reserved for the covariant and absolute derivatives, while the comma is reserved for the normal derivative. So:

$$\boxed{\lambda_{;c}^a = \frac{D\lambda^a}{dx^c} = \frac{\partial \lambda^a}{\partial x^c} + \Gamma_{bc}^a \lambda^b = \partial_c \lambda^a + \Gamma_{bc}^a \lambda^b}$$

and

$$\boxed{\lambda_{,c}^a = \frac{\partial \lambda^a}{\partial x^c} = \partial_c \lambda^a}$$

why do we do this? Because by writing the covariant derivatives in terms of indices, it is easier for us to work with it as a tensor. Recall that λ^a is a type (1,0) tensor. We have also shown that the covariant derivative of a covariant vector is a type (1,1) tensor. By writing the new derivative as $\lambda_{;c}^a$, it is more obvious to us that it is a type (1,1) tensor.

With this notation, we can write the covariant derivative of $\tau_{b;c}^a$ as:

$$\tau_{b;c}^a = \partial_c \tau_b^a + \Gamma_{dc}^a \tau_b^d - \Gamma_{bc}^d \tau_d^a$$

Example 6.8. Show that the metric tensor is covariantly constant, which means that $g_{ab;c} = 0$.

We can start with

$$\Gamma_{bc}^a = \frac{1}{2} g^{ad} (\partial_b g_{cd} + \Gamma_{cd}^a g_{bd} - \partial_d g_{bc})$$

Let us invoke the following quantity, resulted from the lowering of Γ_{bc}^e indices:

$$\begin{aligned} \Gamma_{abc} &= g_{ae} \Gamma_{bc}^e \\ &= \frac{1}{2} g_{ae} g^{ed} (\partial_b g_{cd} + \partial_c g_{bd} - \partial_d g_{bc}) \\ &= \frac{1}{2} (\partial_b g_{ac} + \partial_c g_{ab} - \partial_a g_{bc}). \end{aligned}$$

Swapping a and b so that by a similar reasoning we get

$$\Gamma_{bac} = \frac{1}{2} (\partial_a g_{bc} + \partial_c g_{ba} - \partial_b g_{ac})$$

Combining the previous two results we get

$$\Gamma_{abc} + \Gamma_{bac} = \frac{1}{2} (\partial_c g_{ab} + \partial_c g_{ba}) = \partial_c g_{ab}.$$

So, by definition:

$$\begin{aligned} g_{ab;c} &= \partial_c g_{ab} - \Gamma_{ac}^d g_{db} - \Gamma_{bc}^d g_{ad} \\ &= \Gamma_{abc} + \Gamma_{bac} - \Gamma_{ac}^d g_{bd} - \Gamma_{bc}^d g_{ad} \\ &= \Gamma_{abc} + \Gamma_{bac} - \Gamma_{bac} - \Gamma_{abc} \\ &= 0. \end{aligned}$$

Note that we might have predicted ahead of time that $g_{ab;c} = 0$ by going to a local Lorentz frame so that $g_{\mu\nu;c} = 0$ in 4-dimensional spacetime. There, the metric tensor is simple the Minkowski metric, whose derivatives are zero, resulting in zero-valued Christoffel symbols. Therefore, $g_{\mu\nu;c} = 0$. In fact, following from the fact that tensor equations are covariant.

Proposition 6.1. If a tensor is zero in one frame, then it is zero in all frames.

Let us follow the previous result up with another nice example:

Example 6.9. Show that the Kronecker delta is covariantly constant: $\delta_{b;c}^a = 0$. Then show that $g_{;c}^{ab} = 0$ and $g_{;c}^{ab} = 0$.

This will be left as an exercise for the reader for now.

With the principle of covariance, we now have a prescription for finding physics equations in general relativity:

- Step 1: Write down the physics equations in special relativity
- Change all derivatives to absolute or covariant derivatives, so that the equations turn into tensor equations.
- transform to arbitrary frames where the equations don't change.

Example 6.10. Write Newton's 2nd law in free space as a tensor equation:

In special relativity:

$$f^\mu = \frac{dp^\mu}{d\tau}.$$

To make this equation a tensor equation, we replace the derivative with respect to proper time to an absolute derivative with respect to proper time:

$$f^\mu = \frac{Dp^\mu}{d\tau}.$$

Now let us suppose that the particle in consideration is a free particle (subject to nothing but curvature of spacetime due to gravity). We get

$$f^\mu = \frac{Dp^\mu}{d\tau} = m \frac{Du^\mu}{d\tau} = 0.$$

Therefore,

$$\frac{du^\mu}{d\tau} + \Gamma_{\lambda\nu}^\mu u^\lambda \dot{x}^\nu = 0.$$

But by definition:

$$u^\mu = \frac{dX^\mu}{d\tau},$$

we get

$$\frac{d^2 X^\mu}{d\tau^2} + \Gamma_{\nu\lambda}^\mu \frac{dX^\nu}{d\tau} \frac{dX^\lambda}{d\tau} = 0,$$

which is the geodesic equation, as expected.

6.8 Newtonian limit

6.8.1 Overview

In Newtonian physics, gravity is a force:

$$\vec{F} = -\frac{GMm}{r^2} \hat{r},$$

and the equation of motion is:

$$\vec{F} = \frac{d^2 \vec{X}}{dt^2}.$$

In general relativity, however, the equation of motion is the geodesic equation:

$$\frac{d^2 X^\mu}{d\tau^2} + \Gamma_{\nu\lambda}^\mu \frac{dX^\nu}{d\tau} \frac{dX^\lambda}{d\tau} = 0.$$

Now, in some limit, these equations have to match up. Indeed, there is a quantity in the metric tensor $g_{\mu\nu}$ that links up with a quantity in Newtonian physics. This quantity is called the gravitational potential V . We can think of V by analogy to the electromagnetic potential:

$$V = \frac{U}{q} = -\frac{1}{q} \int \frac{1}{4\pi\epsilon_0} \frac{Qq}{r^2} = \frac{1}{4\pi\epsilon_0} \frac{q}{r}.$$

We can define a similar quantity for gravity, called the gravitational potential:

$$V = \frac{U}{m} = -\frac{1}{m} \frac{GMm}{r} = -\frac{GM}{r}.$$

We can find the relationship between gravitational potential and force:

$$\vec{F} = -m\vec{\nabla}V.$$

So Newton's second law of motion becomes:

$$\vec{F} = m\vec{a} = m\frac{d^2 \vec{X}}{dt^2} = -m\vec{\nabla}V.$$

Therefore,

$$\frac{d^2 \vec{X}}{dt^2} = \vec{\nabla}V.$$

Let us write the above equation in terms of indices:

$$\frac{d^2 X^i}{dt^2} = -\partial_i V.$$

But in order for this equation work with relativistic theory (where we work with four instead of three indices), we "fix" with a Kronecker delta δ^{ij} :

$$\boxed{\frac{d^2 X^i}{dt^2} = -\delta^{ij} \partial_j V}$$

We might wonder how this equation is compatible with the geodesic equation in the non-relativistic limit. We shall explore the answers in the following subsections:

6.8.2 Weak limit of General Relativity

The effects of gravity near Earth or the Sun are weak. Therefore only a slight curvature is expected, which means we can approximate the metric tensor as a Minkowski metric plus a small correction term:

$$g_{\mu\nu} \approx \eta_{\mu\nu} + h_{\mu\nu},$$

where $|h_{\mu\nu}| \ll 1$. In the Newtonian limit, spacetime is almost Minkowski-like (flat). Keeping only the first terms in $h_{\mu\nu}$, we can show the following example that

$$g^{\mu\nu} \approx \eta^{\mu\nu} - h^{\mu\nu},$$

where

$$h^{\mu\nu} = \eta^{\mu\alpha} \eta^{\nu\beta} h_{\alpha\beta}.$$

Example 6.11. Show that $g^{\mu\nu} \approx \eta^{\mu\nu} - h^{\mu\nu}$ where $h^{\mu\nu} = \eta^{\mu\alpha} \eta^{\nu\beta} h_{\alpha\beta}$. Hint: verify that $g^{\mu\nu} g_{\nu\sigma} \approx \delta_{\sigma}^{\mu}$ to first order, assuming that products of h terms are small enough to be taken to 0.

At this point this example is left to the reader.

Once we have these results, we can find the Christoffel symbols $\Gamma_{\nu\sigma}^{\mu}$ in terms of the correction metric h to first order:

$$\Gamma_{\nu\sigma}^{\mu} = \frac{1}{2} \eta^{\mu\rho} (\partial_{\nu} h_{\sigma\rho} + \partial_{\sigma} h_{\nu\rho} - \partial_{\rho} h_{\nu\sigma})$$

Let us use this Christoffel symbol in the geodesic equation, but before we start, we should make a few approximations. First, we want a result in the non-relativistic (slow) limit, so

$$\frac{dX^0}{d\tau} \gg \frac{dX^i}{d\tau}.$$

This follows since

$$\frac{dX^0}{d\tau} = \frac{d}{d\tau}(ct) = c \frac{dt}{d\tau},$$

whereas

$$\frac{dX^i}{d\tau} = \frac{dX^i}{dt} \frac{dt}{d\tau} \ll c \frac{dt}{d\tau}.$$

With this result, we can ignore the terms $dX^i/d\tau$ in summing compared to $dX^0/d\tau$. Now we can start:

$$\frac{d^2 X^{\mu}}{d\tau^2} + \Gamma_{\nu\sigma}^{\mu} \frac{dX^{\nu}}{d\tau} \frac{dX^{\sigma}}{d\tau} \approx \frac{d^2 X^{\mu}}{d\tau^2} + \Gamma_{00}^{\mu} \frac{dX^0}{d\tau} \frac{dX^0}{d\tau} \approx 0.$$

Second, we assume static gravitational field (not changing in time - assuming stationary Earth/Sun), i.e. $\partial_0 h_{\sigma 0} = 0$:

$$\begin{aligned}\Gamma_{00}^\mu &= \frac{1}{2}\eta^{\mu\sigma}(\partial_0 h_{\sigma 0} + \partial_0 h_{\sigma 0} - \partial_\sigma h_{00}) \\ &\approx -\frac{1}{2}\eta^{\mu\sigma}\partial_\sigma h_{00}.\end{aligned}$$

So, the geodesic equation becomes:

$$\begin{aligned}\frac{d^2 X^\mu}{d\tau^2} &\approx \frac{1}{2}\eta^{\mu\sigma}\partial_\sigma h_{00}\left(\frac{dX^0}{d\tau}\right)^2 \\ &= \frac{1}{2}\eta^{\mu\sigma}\partial_\sigma h_{00}c^2\left(\frac{dt}{d\tau}\right)^2.\end{aligned}$$

Let $\mu = 0$.

$$\frac{d^2 X^0}{d\tau^2} = c^2 \frac{d^2 t}{d\tau^2} \approx \left(\frac{1}{2}\eta^{0\sigma}\partial_\sigma h_{00}\right)c^2\left(\frac{dt}{d\tau}\right)^2$$

The only value for σ so that $\eta^{\sigma 0}$ is nonzero is $\sigma = 0$. But because $\partial_0 h_{00} = 0$ in the static limit,

$$\frac{1}{c^2} \frac{d^2 X^0}{d\tau^2} = \frac{d^2 t}{d\tau^2} = 0.$$

Therefore, $dt/d\tau$ has no τ dependence.

Let $\mu = i$,

$$\frac{d^2 X^i}{d\tau^2} \approx \frac{1}{2}\eta^{i\sigma}(\partial_\sigma h_{00})c^2\left(\frac{dt}{d\tau}\right)^2.$$

Applying the chain rule:

$$\begin{aligned}\frac{d^2 X^i}{d\tau^2} &= \frac{d}{d\tau}\left(\frac{dX^i}{d\tau}\right) \\ &= \frac{d}{d\tau}\left(\frac{dX^i}{dt}\frac{dt}{d\tau}\right) \\ &= \frac{d^2 X^i}{dt d\tau}\frac{dt}{d\tau} + \frac{dX^i}{dt}\frac{d^2 t}{d\tau^2} \\ &= \frac{dt}{d\tau}\frac{d}{d\tau}\frac{dX^i}{dt} \\ &= \frac{dt}{d\tau}\frac{d}{dt}\frac{dX^i}{dt}\frac{dt}{d\tau} \\ &= \left(\frac{dt}{d\tau}\right)^2 \frac{d^2 X^i}{dt^2}.\end{aligned}$$

So,

$$\frac{d^2 X^i}{dt^2} = \frac{c^2}{2} \eta^{i\sigma} (\partial_\sigma h_{00}).$$

Now note that this equation indicates a summing of indices. If $\sigma = 0$, then $\eta^{i\sigma} = 0$. While if $\sigma = j$, then $\eta^{ij} = -1 = -\delta^{ij}$ if $i = j$. So,

$$\frac{d^2 X^i}{dt^2} \approx -\frac{c^2}{2} \delta^{ij} (\partial_j h_{00}).$$

Now, we compare this with the Newtonian equation:

$$\frac{d^2 X^i}{dt^2} = -\delta^{ij} \partial_j V.$$

In order for general relativity to go back to Newtonian physics, there must be the following correspondence in the limit:

$$V \approx \frac{c^2}{2} h_{00} + \text{Constant}.$$

Now if we “turn off” gravity, then $h_{\mu\nu} = 0$, since spacetime is completely Minkowski, which means $V = 0$. So, Constant = 0. We get:

$$\boxed{h_{00} = \frac{2V}{c^2}}$$

Now, since

$$g_{00} \approx \eta_{00} + h_{00} = 1 + h_{00},$$

we finally get the correspondence between general relativity and Newtonian physics.

$$\boxed{g_{00} = 1 + \frac{2V}{c^2}}$$

A more careful analysis shows

$$\frac{d\tau}{dt} = \sqrt{1 + h_{00}},$$

which follows from the fact that

$$c d\tau^2 = g_{\mu\nu} dX^\mu dX^\nu$$

is independent of τ in the static and weak limit. We notice that this results in a new type of time dilation which we will explore later when we are close to

discussing cosmology. Basically, the exact solution for a spherical mass M in general relativity is the Schwarzschild solution:

$$g_{\mu\nu} = \begin{pmatrix} 1 - \frac{2GM}{c^2 r} & 0 & 0 & 0 \\ 0 & -\left(1 - \frac{2GM}{c^2 r}\right)^{-1} & 0 & 0 \\ 0 & 0 & -r^2 & 0 \\ 0 & 0 & 0 & -r^2 \sin^2 \theta \end{pmatrix}.$$

Note that

$$g_{00} = 1 + \frac{2V}{c^2},$$

as predicted by our workings so far.

7 Einstein's field equations

We have been assuming that we know the metric tensor and have looked at physics in curved spaces. But we have yet to develop a method to obtain the metric from a given distribution of mass and energy. In this section, we will look at the Einstein equations - our prescription to finding the metric tensor given a distribution of mass and energy. The Einstein equations took Einstein 8 years to develop. The Einstein equations are a set of coupled, non-linear partial differential equations that we can solve for the metric tensor $g_{\mu\nu}$.

As a little teaser, let us look at the form of the Einstein equations:

$$R^{\mu\nu} - \frac{1}{2}Rg^{\mu\nu} = -\frac{8\pi G}{c^4}T^{\mu\nu}$$

where

- $T^{\mu\nu}$ is the energy-momentum stress tensor, which represents the density of energy, mass, and momentum, which together is the source of gravity or the curvature of spacetime.
- $R^{\mu\nu}$ is the Ricci tensor. This is actually a contraction of the Riemann curvature tensor $R^\mu{}_{\nu\lambda\sigma}$:

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

- R is the curvature scalar, which is the contraction of the Ricci tensor $R_{\mu\nu}$:

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

We will eventually show that the Riemann curvature tensor $R^\mu{}_{\nu\sigma\lambda}$ is a function of the metric tensor $g_{\mu\nu}$ and its derivatives.

As a little aside: When Einstein looked at the solutions for a gas of cosmic matter, he found an evolving solution which suggested an expanding/contracting universe. However, because Einstein thought the universe has to be static (as he was doing this before Hubble's discover in 1929 that the universe is expanding), Einstein added an extra term Λ to his equations:

$$R^{\mu\nu} - \frac{1}{2}g^{\mu\nu}R + \Lambda g^{\mu\nu} = \frac{8\pi G}{c^4}T^{\mu\nu}.$$

Λ is the **cosmological constant**. This extra term accounts for a cosmic source of energy density that is associated with the vacuum (which we today refer to as **dark energy**). After Hubble's discovery, though, Einstein set Λ to 0, calling it the "greatest blunder" of his life, and for decades, all cosmology uses $\Lambda = 0$ until the 1990s when it was discovered that the universe is not only expanding

but also the expansion is accelerating. This event brought back Λ . Today, almost all cosmological models include the cosmological constant or some form of “dark energy.”

In the later sections, we will study cosmology with and without the cosmological constant Λ . Our plan is to look at $T^{\mu\nu}$, $R^{\mu\nu\lambda\sigma}$, $R_{\mu\nu}$, R , etc. and retrace some of Einstein’s steps with which he came up with his solutions.

We should also note that the Einstein equations are very difficult to solve partly because they are non-linear: Gravitational fields carry energy, which in turn affects themselves, unlike in electromagnetism where we get linear equations whose solutions obey the principle of superposition because electromagnetic waves don’t carry charge and don’t interact with each other. Now, we are not going to attempt to solve the Einstein equations (that would be way beyond the scope of this Quick Guide). Instead, we will study two well-known solutions to the Einstein equations:

- The Schwarzschild solution, which gives the metric tensor outside a spherical static mass M such as the Sun or the Earth or a black hole.
- The Friedmann-Robertson-Walker (FRW) solution, which gives the metric tensor for a homogeneous and spatially isotropic universe with $\Lambda = 0$ or $\Lambda \neq 0$. Note that the FRW solution with $\Lambda \neq 0$ is the current best cosmological model.

Now, let us find out how Einstein was guided to find what we now call the Einstein equations. In Newtonian limit, for a point particle, we know that

$$\vec{F} = -m\vec{\nabla}V,$$

where

$$V = -\frac{GM}{r}.$$

What about for a mass density of ρ ? Let us first define the mass density ρ as:

$$\boxed{\vec{\nabla}^2 V = 4\pi G\rho}$$

This is called the **Poisson equation**. How we show this is true? We can do this by analogy (again) to electromagnetism. In electromagnetism, the charge density is found through the first of four Maxwell’s equations:

$$\rho = \epsilon_0 \vec{\nabla} \cdot \vec{E},$$

where \vec{E} can be expressed in terms of the electric potential:

$$\vec{E} = -\vec{\nabla}V.$$

This simply gives:

$$\vec{\nabla}^2 V = -\frac{\rho}{\epsilon_0}.$$

Einstein indeed used this analogy as a guide.

7.1 The stress-energy tensor $T^{\mu\nu}$

As we have said before, $T^{\mu\nu}$ is the energy-momentum stress tensor. For a given distribution of matter,

$$\rho = \frac{dM}{dV}$$

is the mass density, while ρc^2 gives the mass-energy density. We also know that energy and momentum couple relativistically, so we might be wondering what the momentum-type density (that goes with mass density) is. The answer, it turns out, to be pressure, p , (force per area). So, in general relativity, the pressure p acts as a source of energy-momentum density. But note that p is not a vector, so we expect p and ρc^2 to be the matrix elements of the tensor $T^{\mu\nu}$. Also, since $g^{\mu\nu} = g^{\nu\mu}$ (symmetry), we expect a similar symmetry $T^{\mu\nu} = T^{\nu\mu}$. Therefore, for a simple gas of particles in a rest frame in flat spacetime, we expect $T^{\mu\nu}$ to have the following form:

$$T^{\mu\nu} = \begin{pmatrix} \rho c^2 & 0 & 0 & 0 \\ 0 & p & 0 & 0 \\ 0 & 0 & p & 0 \\ 0 & 0 & 0 & p \end{pmatrix}.$$

But to put $T^{\mu\nu}$ into covariant form that allows for moving matter we need to take into account in the world velocity u^μ . We can show that $T^{\mu\nu}$ has the following form:

$$T^{\mu\nu} = \left(\rho + \frac{p}{c^2} \right) u^\mu u^\nu - p \eta^{\mu\nu}$$

where $u^\mu = (\gamma c, \gamma \vec{v})$ for moving matter, and $g^{\mu\nu} = \eta^{\mu\nu}$ in flat spacetime. Einstein knew this was the quantity to use because it obeys the conservation law:

$$T^{\mu\nu}_{;\mu} = 0 \equiv \partial_\mu T^{\mu\nu} = 0$$

This gives two well-known equations in fluid dynamics: the **continuity equation**:

$$\frac{\partial \rho}{\partial t} + \vec{\nabla} \cdot \rho \vec{v} = 0,$$

and Euler's equation:

$$\rho \left(\frac{\partial}{\partial t} + \vec{v} \cdot \vec{\nabla} \right) \vec{v} = -\vec{\nabla} p.$$

Now, we can also have energy density from electromagnetism because electric and magnetic fields carry energy and momentum. So, we define a stress tensor for them as:

$$T_{\text{EM}}^{\mu\nu} = F^\mu_\lambda F^{\nu\lambda} + \frac{1}{4} \eta^{\mu\nu} F_{\alpha\beta} F^{\alpha\beta},$$

where $F^{\mu\nu}$ is the electromagnetic stress tensor. (For now, we will not show how this definition is motivated). With this definition, we can define the total energy-momentum stress tensor as the element-wise sum of the stress tensors:

$$T^{\mu\nu} = T_{\text{matter}}^{\mu\nu} + T_{\text{EM}}^{\mu\nu} + \dots$$

Lastly, to make the equations covariant, we simply replace η with g and the comma derivative with semi-colon derivatives. This gives the energy-momentum stress tensor of matter in general relativity as

$$T_{\text{matter}}^{\mu\nu} = \left(\rho + \frac{p}{c^2} \right) u^\mu u^\nu - p g^{\mu\nu}.$$

And the conservation law:

$$T^{\mu\nu}{}_{;\mu} = 0.$$

Next, how do we find an equation that lets us solve for the metric tensor $g_{\mu\nu}$ given a distribution of matter $T^{\mu\nu}$. An obvious first guess is:

$$g^{\mu\nu} = k T^{\mu\nu}.$$

If this is the case, then $g^{\mu\nu}$ inherits all properties of $T^{\mu\nu}$. This is good, but it doesn't give the Poisson equation. So, we want to look for a relationship that involves the Christoffel symbols. But remember that $\Gamma_{\mu\nu}^\lambda$ is NOT a tensor. Also, recall that $\Gamma_{\lambda\nu}^\mu \neq 0$ does not mean spacetime is curved. These two reminders motivate us (or rather, Riemann) to find a quantity that describes the curvature of spacetime. This quantity is called the Riemann curvature tensor:

$$R^\mu{}_{\nu\lambda\sigma} = \partial_\lambda \Gamma_{\nu\sigma}^\mu - \partial_\sigma \Gamma_{\nu\lambda}^\mu + \Gamma_{\nu\sigma}^\rho \Gamma_{\rho\lambda}^\mu - \Gamma_{\nu\lambda}^\rho \Gamma_{\rho\sigma}^\mu$$

Spacetime is flat if the Riemann curvature tensor is zero at all points. If the Riemann curvature tensor is nonzero at any point in spacetime, then spacetime is curved.

7.2 Riemann curvature tensor $R^\lambda{}_{\mu\nu\sigma}$

The Riemann curvature tensor is obtained by doing repeated covariant differentiation. Notice, though, that we are concerned with covariant differentiation rather than usual derivatives, which obey "the equality of mixed partials:"

$$\frac{\partial^2}{\partial X^\mu \partial X^\nu} = \frac{\partial^2}{\partial X^\nu \partial X^\mu}$$

As we have learned, the equality of mixed partials does not hold when is curvature. Suppose that λ^μ is a covariant vector:

$$\lambda_{\nu;\sigma} = \partial_\sigma \lambda_\nu - \Gamma_{\nu\sigma}^\mu \lambda_\mu,$$

and

$$\lambda_{\nu;\sigma\lambda} = (\lambda_{\nu;\sigma})_{;\lambda} = (\partial_\sigma \lambda_\nu - \Gamma_{\nu\sigma}^\mu \lambda_\mu)_{;\lambda}.$$

We can easily show that the equality of mixed partials no longer hold in general:

$$\lambda_{\nu;\lambda\sigma} \neq \lambda_{\nu;\sigma\lambda}.$$

However, it is even more interesting that:

$$\lambda_{\nu;\lambda\sigma} - \lambda_{\nu;\sigma\lambda} = R^\mu_{\nu\lambda\sigma} \lambda_\mu$$

where if the Riemann curvature tensor is nonzero then the covariant second-derivatives (of different order of differentiation) of λ_μ are not equal.

Example 7.1. Prove the following cyclic identity: $R^\mu_{\nu\lambda\sigma} + R^\mu_{\lambda\sigma\nu} + R^\mu_{\sigma\nu\lambda} = 0$.

For now the exercise is left to the reader.

We observe that the Riemann curvature tensor has four indices, each with four values 0 to 3. We might be expecting $4^4 = 256$ indices, but we will show later on that they are not all mutually independent. Apart from the cyclic identity above, the Riemann curvature tensor has a number of interesting properties. We shall look at some important ones.

If we lower an index:

$$R_{\mu\nu\lambda\sigma} = g_{\mu\sigma} R^\sigma_{\nu\lambda\sigma},$$

we can show that:

$$\begin{aligned} R_{\mu\nu\lambda\sigma} &= -R_{\nu\mu\lambda\sigma} \\ R_{\mu\nu\lambda\sigma} &= -R_{\mu\nu\sigma\lambda} \\ R_{\mu\nu\lambda\sigma} &= R_{\lambda\sigma\mu\nu}. \end{aligned}$$

All of these properties follow from the definition in terms of the Christoffel symbols $\Gamma_{\nu\sigma}^\mu$ and the metric tensor $g_{\mu\nu}$. With all these relations, we can show that there are only 20 independent components in the Riemann curvature tensor $R^\mu_{\nu\lambda\sigma}$.

Next, what if we contract an index of the Riemann curvature tensor?

$$R^\mu_{\mu\lambda\sigma} = g^{\mu\rho} R_{\rho\mu\lambda\sigma} = -g^{\mu\rho} R_{\mu\rho\lambda\sigma} = -R^\rho_{\rho\lambda\sigma} = -R^\mu_{\mu\lambda\sigma}.$$

This simply says that the contraction vanishes:

$$R^\mu_{\mu\lambda\sigma} = 0$$

But there are contractions that don't vanish. For example, the Ricci tensor:

$$R_{\mu\nu} = R^\sigma_{\mu\lambda\sigma}.$$

Example 7.2. Show that the Ricci tensor is symmetric: $R_{\mu\nu} = R_{\nu\mu}$.

We start with the cyclic identity:

$$R^\mu{}_{\nu\lambda\sigma} + R^\mu{}_{\lambda\sigma\nu} + R^\mu{}_{\sigma\nu\lambda} = 0.$$

We then contract σ with μ (or let $\sigma = \mu$):

$$\begin{aligned} R^\mu{}_{\nu\lambda\mu} + R^\mu{}_{\lambda\mu\nu} + R^\mu{}_{\mu\nu\lambda} &= 0 \\ R^\mu{}_{\nu\lambda\mu} + R^\mu{}_{\lambda\mu\nu} &= 0 \\ R^\mu{}_{\nu\lambda\mu} - R^\mu{}_{\lambda\mu\nu} &= 0 \\ R_{\mu\nu} &= R_{\nu\mu}. \end{aligned}$$

The significance of this example is that it shows $R_{\mu\nu}$ has only 10 independent components just like $g_{\mu\nu}$.

7.3 The Einstein equations

Before we get into the Einstein equations, we define the **curvature scalar**:

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

7.4 Schwarzschild solution

8 Predictions and tests of general relativity

8.1 Gravitational redshift

8.2 Radar time-delay experiments

8.3 Black Holes

9 Cosmology

9.1 Large-scale geometry of the universe

9.1.1 Cosmological principle

9.1.2 Robertson-Walker (flat, open, closed) geometries

9.1.3 Expansion of the universe

9.1.4 Distances and speeds

9.1.5 Redshifts

9.2 Dynamical evolution of the universe

9.2.1 The Friedmann equations

9.2.2 The cosmological constant Λ

9.2.3 Equations of state

9.2.4 A matter-dominated universe ($\Lambda = 0$) [Friedmann models]

9.2.5 A flat, matter-dominated universe ($\Lambda = 0$) [old favorite model]

9.3 Observational cosmology

9.3.1 Hubble law

9.3.2 Acceleration of the universe

9.3.3 Matter densities and dark matter

9.3.4 The flatness and horizon problems

9.3.5 Cosmic Microwave Background (CMB) anisotropy

9.4 Modern cosmology

9.4.1 Inflation

9.4.2 Dark energy (The cosmological constant problem)

9.4.3 Concordance model [new favorite model]

9.4.4 Open questions