

PHYS 514: General Relativity Lecture Notes

Shereen Elaidi

Contents

1 Lecture 1: 7 January 2020	1
2 Lecture 2: 12 January 2021	2
3 Lecture 3: 14 January 2021	3
3.1 Special Relativity and the Equivalence Principle	3
3.2 Why GR is a simple modification of SR (Equivalence Principle)	5
4 Lecture 4: 19 January 2020	6
4.1 Manifolds	6
5 Lecture 5: 21 January 2021	8
5.1 Tensors	8
6 Lecture 6: Tensors 26 January 2021	10

1 Lecture 1: 7 January 2020

What is GR? GR is the modern theory of classical gravity that incorporates the effects of special relativity. The basic idea is the following: gravity is a force. Most forces are described by fields. *What are fields?* We have the following examples of fields:

- (1) The Newtonian gravitational Potential, Φ .
- (2) For electromagnetism, we have the electromagnetic fields E and B .

We call such a theory a **field theory**. A classical field theory has two ingredients that make up a theory of forces described by fields:

- (1) **A field equation:** an equation of motion which tells us exactly how the field is determined by some set of forces. This is typically a 2nd order ODE. For example:
 - (1) The field equation for Newtonian gravity is $\nabla^2\Phi = 4\pi G\rho$.
 - (2) The field equations for E&M are Maxwell's equations.
- (2) **Force Law:** an equation which determines how objects move in the presence of a field. For example:
 - (1) For Newtonian gravity, this is $F = ma = m\nabla\Phi$.
 - (2) For E&M, this is the Lorentz force law $F = q(E + v \times B)$.

The fields are functions of points in space-time. We write them as $\Phi(t, x)$ for the gravitational field or $E(t, x)$ for the electric field. The force law tells us how the objects move, i.e., how they deviate from a straight line ($a = 0$). The modern perspective of GR is the following: it is not a field theory; it is something totally different— gravity is not due to a field Φ that is a function of space-time, but is instead a feature of space-time itself. Hence, we replace the Φ with a “metric tensor” which describes the curvature / geometry of space-time.

The field equation of GR determines how space-time curves in the presence of matter or energy. This field equation is a differential equation for the metric of space-time that helps us determine how space-time curves in the presence of matter.

The force law is the geodesic equation, which tells us how objects move through curved space-time. In short, this means that objects travel on **geodesics**, which are as straight as possible given the curvature of space-time. For example, on a plane, the shortest way to travel between two points is simply a line. However, for a sphere, the curve that minimises distance is the arc of a great circle.

2 Lecture 2: 12 January 2021

Space time is a set of points at which events could take place, which can be parameterized in a smooth manner by a coordinate system, such as cartesian coordinates (t, x) . We call points in space-time **events**. Physics should be independent of the coordinate system used.

In Newtonian physics, for two points (t_1, x_1) and (t_2, x_2) , we can define the time-separation and spatial separations:

$$\begin{aligned}\Delta t &= t_2 - t_1, \\ \Delta x &= \sqrt{(x_2 - x_1)^2}.\end{aligned}$$

These two quantities have independent meaning and make sense in Newtonian physics (this is called the principle of general covariance). In SR, there is no separate notion of time and space (time separation/distance) between two events. In other words, Δt and Δx do not make sense independently. These depend on the frame of reference in which they are measured. But, there is a notion of an “invariant interval” between two events. Given two events,

$$\begin{aligned}\Delta s^2 &= -c^2 \Delta t^2 + \Delta x^2 \\ &= -\Delta t^2 + \Delta x^2 \text{ (units where } c = 1 \text{)}\end{aligned}$$

When we talk about space-time, it’s useful to draw pictures. We’ll plot pictures in space-time using a **space-time diagram**, with time running vertically. Curves of a fixed Δs trace out hyperbolae. We have that light-rays travel on 45 degree lines. We have three cases:

- (1) $\Delta s^2 = 0$: these two points can be connected by a light ray. These pairs of events are called **null-separated** or **light-like separated**.
- (2) $\Delta s^2 < 0$: these points/events can be connected with a trajectory with a velocity less than c . We call such events **time-like separated**.
- (3) $\Delta s^2 > 0$: these are points which cannot be connected as above. We call such events **space-like separated**.

Claim: all of special relativity can be reduced to the following statement:

For two time-like separated events, the proper time $\Delta\tau$ measured by an observer moving at a constant velocity between the events is given by:

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

This reproduces various phenomena. For example, consider **time-dilation**. Using the above, we can deduce that the relative age difference between the two twins would be $\frac{2\Delta\tau}{\Delta t} = \sqrt{1 - v^2}$. We see a similar result for **length contraction**: $\frac{\Delta\tau^2}{t^2} = 1 - v^2$. In our world, we have three spatial coordinates, and so the invariant interval takes the following form:

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

To write this more compactly, we'll introduce a new notation. Let $x^\mu = (t, x)$ where $\mu = 0, 1, 2, 3$. Then, we can write the invariant interval as:

$$(\Delta s)^2 = \sum_{\mu, \nu=0}^3 g_{\mu\nu} \Delta x, \quad (1)$$

where,

$$g_{\mu\nu} = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

That matrix is the “metric” of flat space-time in cartesian coordinates. We can write things even more compactly by introducing **Einstein summation notation**:

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

We read this by knowing that repeated indices are summed. Since μ and ν range from 0 to 3, we will be summing the above from 0 to 3.

You are familiar with Euclidean space \mathbb{R}^2 and \mathbb{R}^3 . Minkowski space is denoted by $\mathbb{R}^{3,1}$; this is just like \mathbb{R}^4 except we have a negative sign in the formula for the distance. This gives us a **Lorentzian space-time**.

3 Lecture 3: 14 January 2021

3.1 Special Relativity and the Equivalence Principle

We left off last time with a reformulation of special relativity as the geometry of space-time. In particular,

The proper time $\Delta\tau$ measured by an observer moving at a constant velocity between two points in space-time is given by the invariant interval Δs :

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

where $\Delta s^2 = -\Delta t^2 + (\Delta x)^2$ in Cartesian coordinates.

Recalling that $\Delta x^\mu = (\Delta t, \Delta x)$, we can write the above using the metric as:

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

where the metric is given by the following matrix:

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

Now, we want to build up to the equivalence principle. This requires us to think about observers who are accelerating (basic idea of the equivalence principle). For us, accelerating means that observers are moving through a path in space-time which is not in a straight line.

A general path through space-time is known as a **world-line**. It is parameterized by four functions. Let λ be our parameter which labels different points on the curve. Then:

$$x^\mu(\lambda) = (t(\lambda), x^1(\lambda), x^2(\lambda), x^3(\lambda)).$$

Q: How would you measure the proper time of an observer moving on this trajectory: $\Delta\tau$?

The proper time $\Delta\tau$ measured on an observer can be obtained by taking the following limiting process: break the curve down into small segments, approximate each segment by straight lines, then apply the formula from the previous lecture to proper time along a straight line. Then, take the limit as the length of each interval goes to zero. This gives us that the proper time along a path is $\sqrt{-(\Delta s)^2}$, where

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

Observe that the above is similar to the arc-length of a curve from calculus. Using the Einstein summation notation, we can write this more compactly as:

$$\Delta s = \int \sqrt{\eta_{\mu\nu} \frac{\partial x^\mu}{\partial \lambda} \frac{\partial x^\nu}{\partial \lambda}} d\lambda. \quad (3)$$

We will write this as $\Delta s = \int ds$, where ds is an infinitesimal interval of space-time. We call it the **line-element of ST in GR**:

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

which we got by cancelling the $d\lambda$ in the above. Now, special relativity is the statement that “line element” is given by

$$ds^2 = \eta_{\mu\nu} dx^\mu dx^\nu \text{ with } \eta_{\mu\nu} = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

GR is a very simple modification of the above statement; GR is the statement that all the effects of gravity are packages in terms of the replacement of $\eta_{\mu\nu}$ by $g_{\mu\nu}(x)$; the components depend on the coordinates themselves. We call this the “metric of curved space-time.” $\eta_{\mu\nu}$ is parameterized by 16 functions of x . Hence, to summarise, every statement of ST goes to GR with the proviso that

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu. \quad (4)$$

In GR, light rays still travel along null trajectories, but the null trajectory is defined as $ds = 0$ rather than $\Delta s = 0$ as previously. Similarly, time-like trajectories (trajectories which massive objects travel along) has $(ds)^2 < 0$ for every point along the trajectory. The proper time measured by an observer who travels along a worldline $x^\mu(\lambda)$ is again given by:

$$\Delta\tau^2 = -\Delta s^2, \quad (5)$$

where $\Delta s = \int ds$ where ds is given by:

$$\Delta s = \int \sqrt{g_{\mu\nu} \frac{dx^\mu}{d\lambda} \frac{dx^\nu}{d\lambda}} d\lambda. \quad (6)$$

With this geometric formulation, general relativity is a nearly trivial modification of special relativity. Just replace $\eta_{\mu\nu}$ with $g_{\mu\nu}$. We still need some dynamical principle (this will be Einstein’s field equations) and we need some mathematical tools. For example, the above gives us time dilation, which then give us phenomena such as red shift in cosmology and event horizons for black holes.

3.2 Why GR is a simple modification of SR (Equivalence Principle)

The equivalence principle starts with a very simple but powerful observation about gravity. In Newtonian gravity, the way that you describe dynamics is by determining force. Let m_i be the inertial mass of an object and let m_g be the gravitational mass of an object. Let Φ be the gravitational potential. Then,

$$\begin{aligned} F &= m_i a \\ &= -m_g \nabla \Phi. \end{aligned}$$

In principle, m_i may not be equal to m_g . But, it is an *experimental fact* that for every object that we know, $m_i = m_g$. You can cancel them out and get the first **equivalence principle**:

$$\boxed{a = -\Delta \Phi} \quad (7)$$

for all objects, independent of their mass (think of the Leaning tower of Pisa experiment or the feather dropping experiments from space). Note that it is possible that we could discover a new form of matter for which this is not true, in which case we would need to modify GR to deal with that. The above implies that in the presence of a gravitational potential, there is a preferred set of trajectories which any freely-falling object would prefer. These trajectories are the solutions of $a = -\Delta \Phi$. We call those trajectories **inertial trajectories**.

Consider a small region of space-time where $\nabla \Phi$ is approximately constant. Call that constant $-a_0$. The basic idea of the **Weak Equivalence Principle** is the following:

In this very small region, the effects of gravity are *indistinguishable* from the effects of being in a constantly accelerating reference frame.

We can see this via a thought experiment. Consider two labs: one situated on Earth, and another attached to the end of a rocket in space which is accelerating with $a_0 = g$. The EOM of an object in the first lab is $\ddot{x} = -g$. The EOM of an object in the second lab would be $x' = x - \frac{1}{2}gt^2$, i.e., $\ddot{x}' = -g$. This means that if you were to do experiments in these labs, there is no way to tell which scenario that you are in, i.e., the laws are the same. We call forces which arise due to acceleration **fictitious forces**.

The equivalence principle is the statement that in a very small area, the effects of gravity are indistinguishable from a fictitious force due to acceleration. In this case, the metric of flat space-time (Minkowski space) will no longer be constant when written in terms of accelerating coordinates, which then gives us that the effects of gravity can be captured by changes in the metric.

Once $\Delta \Phi$ is no longer approximately constant, i.e., when we are no longer in a region of space-time which is small, we cannot just think of gravity in terms of acceleration (fictitious forces). For example, in a very lab due to tidal forces, we can use experiments to tell if there is varying gravitational potential present; this cannot be mimicked by a constantly accelerating reference frame.

Let's conclude by reformulating the EP (Equivalence Principle) a bit:

Einstein Equivalence Principle: in a sufficiently small region of space-time, the laws of physics reduce to those of SR. In other words, space-time *locally* looks like Minkowski space $\mathbb{R}^{3,1}$.

We'll spend the next few weeks unpacking the above statement. There's also a V2 of the EEP:

EEP': for any point x_0^μ , there is a coordinate system such that the line element takes the form of that in SR near x_0^μ . Mathematically:

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

where $O(x - x_0^\mu)$ represents the extent to which gravity is not a fictitious force.

4 Lecture 4: 19 January 2020

4.1 Manifolds

Space-time locally “looks like” Minkowski space $\mathbb{R}^{3,1}$. On larger scales, gravity cannot just be regarded as a fictitious force. It’s in the sense that a sphere locally looks like flat space. Precisely, this is the statement that ST is a four-dimensional pseudo-Riemannian **manifold**. We say that A D -dimensional **manifold** M is a set of points that can be labelled by coordinate systems in a smooth way.

Precisely, a coordinate system is a set of d functions x^i which label different points in some region $U \subseteq M$ in a unique way. Mathematically, this is expressed as:

$$x^i : \mathcal{U} \rightarrow \mathbb{R}^d, \quad i = 1, \dots, D.$$

The region \mathcal{U} which these coordinates label uniquely is called a **patch** or **coordinate patch** of the manifold which is “covered” by the x^i . A **global coordinate system** is one which covers all of M . Generally, there are many coordinate systems which may cover some patch \mathcal{U} of M .

E.g.: x^i and $x^{i'}$ are two different ways of labelling points in \mathcal{U} . Then, we can talk about the coordinate transformation or a change of coordinates between x^i and $x^{i'}$. These functions will be parameterized and C^∞ .

$$x^i(x^{i'}) \text{ or } x^{i'}(x^i).$$

If the coordinates x^i and $x^{i'}$ uniquely, label points in some region \mathcal{U} , then the Jacobian matrix

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

is invertible (each index runs from $1, \dots, D$). This describes the components of a $D \times D$ matrix. If it’s invertible, then we have that $\det\left(\frac{\partial x^i}{\partial x^{j'}}\right) \neq 0$. If the determinant is zero, then one of the coordinate system has become degenerate in some way in the sense that it no longer labels points in a unique way. Hence, the coordinate system becomes singular.

For example, the plane \mathbb{R}^2 describes a two-dimensional manifold. Two coordinate systems on this manifold are the cartesian coordinates (x, y) (global coordinates) and polar coordinates (r, θ) . They are related by a change of coordinates:

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

Then, the coordinate transformations arise given by:

$$x^i(x^{i'}) : x = r \cos \theta, \quad y = r \sin \theta$$

and the inverse transformation is given by:

$$x^{i'}(x^i) : r = \sqrt{x^2 + y^2}, \quad \theta = \arctan(y/x).$$

Hence, the Jacobian matrix is given by:

$$\left(\frac{\partial x^i}{\partial x^{j'}} \right) = \begin{bmatrix} \cos \theta & \sin \theta \\ -r \sin \theta & r \cos \theta \end{bmatrix}$$

We have that $\det(\text{jac}) = r \cos^2 \theta + r \sin^2 \theta = r$, which is non-zero only if $r \neq 0$.

When describing space-time, we usually use a spatial notation for the coordinates:

$$x^\mu, \mu = 0, \dots, 3.$$

Having greek indices mean that we are describing space-time; having Roman indices mean that we are describing a Euclidean manifold. For example:

$x^\mu = (t, x)$ of $\mathbb{R}^{3,1}$ is an example of a coordinate system on a space-time.

Most important sentence of this lecture: our goal is to write the laws of physics in such a way that any physical prediction made by these laws is independent of the choice of coordinate systems that we use to write the law. For example, we had the same laws in Newtonian physics for both polar and cartesian coordinates. We give this goal the following name: **Principle of General Covariance**.

E.g., in previous lectures we met the line element $ds^2 = \eta_{\mu\nu} dx^\mu dx^\nu$. Under some change of coordinates $x^\mu \rightarrow x^{\mu'}(x^\mu)$, we have that the line element transforms by multiplication by the Jacobian matrix:

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

What this is saying is the following: if I displaced ν' by some infinitesimal amount, and asked how much μ changed, then the above equation (basically thanks to the chain rule), will answer that question for us. This becomes clear with an example:

E.g.: set $x = r \cos \theta, y = r \sin \theta$. Then,

$$\begin{aligned} dx &= \cos \theta dr - r \sin \theta d\theta, \\ dy &= \sin \theta dr + r \cos \theta d\theta. \end{aligned}$$

How does the line element transform? We apply the formula:

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

where

$$g_{\mu'\nu'} = \eta_{\mu\nu} \frac{\partial x^\mu}{\partial x^{\mu'}} \frac{\partial x^\nu}{\partial x^{\nu'}}.$$

In other words, the metric transforms as:

$$\eta_{\mu\nu} \rightarrow g_{\mu'\nu'} = \eta_{\mu\nu} \frac{\partial x^\mu}{\partial x^{\mu'}} \frac{\partial x^\nu}{\partial x^{\nu'}}.$$

E.g.: the line element of flat space \mathbb{R}^2 is the thing that we use to measure distances:

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

In polar coordinates, this becomes:

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

5 Lecture 5: 21 January 2021

5.1 Tensors

Our goal is to formulate the laws of physics such that things are independent of the coordinate systems in which we use to measure them. This goal is called the **Principle of Covariance**. This goal underlies everything that we will be discussing today. To tackle this goal, we will need to understand how quantities transform under coordinate transformations.

To that end, consider a D -dimensional manifold M . A patch M can be described by multiple different coordinate systems; say, x^μ and $x^{\mu'}$. The simplest object that can transform is a **scalar/scalar function**.

A **scalar function** f is an assignment of a number to each point on a manifold M which is *independent* of the coordinates. For example:

$$f(x^\mu) \text{ in } x^\mu \text{ coordinates.}$$

$$f(x^{\mu'}) \text{ in } x^{\mu'} \text{ coordinates} = f(x^{\mu'}(x^\mu)) \text{ in } x^\mu \text{ coordinates.}$$

We say that f is invariant under the change of coordinates $x^\mu \rightarrow x^{\mu'} = x^{\mu'}(x^\mu)$. A scalar function is called a **(0,0)-tensor**.

A **vector** is a set of D functions V^μ (in each coordinate system) which transforms under changes of coordinates in a particular way:

$$V^\mu(x^\mu) \text{ in } x^\mu \text{ coordinates.}$$

$$V^{\mu'}(x^{\mu'}) = \frac{\partial x^{\mu'}}{\partial x^\mu} V^\mu(x^{\mu'}(x^\mu)).$$

This is known as a **contravariant** vector or a $(1,0)$ -tensor. This is a generalisation of the usual idea of a vector in multivariable calculus. For example, consider \mathbb{R}^3 , and let v be a vector. We can represent v in a specific coordinate system with the following components: $v = (v^1, v^2, v^3)$. If we were to rotate the axes, then we can express the rotation with a rotation matrix R :

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = R \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

The components of v in this new coordinate system would accordingly change:

$$\begin{bmatrix} v^{1'} \\ v^{2'} \\ v^{3'} \end{bmatrix} = \begin{bmatrix} R_1^1 & R_1^2 & R_1^3 \\ \vdots & \ddots & \vdots \\ R_3^1 & \dots & R_3^3 \end{bmatrix} \begin{bmatrix} v^1 \\ v^2 \\ v^3 \end{bmatrix}$$

The *components* of the vector will transform by the multiplication of the transformation matrix. To see how this exactly works out, let v^i be the components of v in the unprimed coordinate frame, and let $v^{i'}$ be the components in the un-primed frame. Using index multiplication, we will discover that Einstein summation notation is basically repackaging matrix multiplication:

$$v^{i'} = R_j^{i'} v^j,$$

that is, R is simply the Jacobian:

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

So, our definition of a vector is simply an object that obeys the following transformation rule: $v^{\mu'} = \frac{\partial x^{\mu'}}{\partial x^{\mu}} v^{\mu}$; it's just a generalisation from vector calculus which accommodates more interesting and non-linear coordinate transformations. This is interesting to us if we think about Einstein's Equivalence Principle. Even tho the components of our vector,

$$v^{\mu} = \begin{bmatrix} v^0 \\ \vdots \\ v^3 \end{bmatrix}$$

will look different in different coordinate systems, a vector should be regarded as a **geometric object** with *coordinate-independent meaning*. A typical example is as follows: consider a world line $x^{\mu}(\lambda)$ which is moving through the manifold. The points on this curve are parameterised by some variable λ . We can then define the vector $v^{\mu} := \frac{\partial x^{\mu}}{\partial \lambda}$; this is the tangent vector to the world-line.

Why is this a vector in the sense of tensor calculus? This is a vector because of the chain rule; the chain rule says that in the new coordinates $x^{\mu'}(x^{\mu})$:

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

where $\left(\frac{\partial x^{\mu'}}{\partial x^{\mu}} \right)$ is the Jacobian matrix. We can consider the set of all tangent vectors – the **tangent space**, denoted $T_p(M)$ – is the set of all tangent vectors attached to a point p of M . This is a vector space. A tangent vector is an example of a vector that is defined for each point on a curve. A vector attached to each point in M is called a **vector field**; in this class, if we talk about a vector, we mean a VECTOR FIELD.

We can also use our vectors to define differential operators. Given a scalar f and a vector $v^{\mu}(x^{\mu})$, we can define the **differential operator** v by:

$$vf := v^{\mu} \frac{\partial}{\partial x^{\mu}} f = v^{\mu} \partial_{\mu} f = v^{\mu} f_{,\mu}.$$

This is precisely the directional derivative in the direction v . For this to be well-defined, vf must be a scalar. Indeed, we have that it is a scalar by the chain rule.

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

In particular, by the chain rule,

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

Since the product of the two Jacobians,

$$J = \begin{bmatrix} \frac{\partial x^{1'}}{\partial x^1} & \cdots & \frac{\partial x^{D'}}{\partial x^1} \\ \vdots & & \vdots \\ \frac{\partial x^{1'}}{\partial x^D} & \cdots & \frac{\partial x^{D'}}{\partial x^D} \end{bmatrix}$$

and

$$\hat{J} = \begin{bmatrix} \frac{\partial x^1}{\partial x^{1'}} & \cdots & \frac{\partial x^D}{\partial x^{1'}} \\ \vdots & & \vdots \\ \frac{\partial x^1}{\partial x^{D'}} & \cdots & \frac{\partial x^D}{\partial x^{D'}} \end{bmatrix}$$

gives us the identity, that means:

$$vf = v^\mu \delta_\mu^\nu \frac{\partial f}{\partial v} = v^\mu \frac{\partial}{\partial x^\mu} f.$$

This verifies that the differential operator $v = v^\mu \partial_\mu$ takes scalars to scalars. We call this the **directional derivative in the direction of v** . If we had some scalar function,

$$f(x^\mu(\lambda)) = f(\lambda),$$

then we can interpret the directional derivative as:

$$\frac{df}{d\lambda} = \frac{\partial x^\mu}{\partial \lambda} \frac{\partial f}{\partial x^\mu} = v^\mu \frac{\partial}{\partial x^\mu} f.$$

The geometric interpretation of this is as follows: this operator takes a derivative along the curve to which v is a tangent vector. Sometimes, we think of this ∂_μ as a “basis” for the set of vector in the same sense that the e_i form a standard basis of \mathbb{R}^n :

$$v = v^\mu \partial_\mu$$

6 Lecture 6: Tensors 26 January 2021

A tensor is a generalisation to the manifold the idea of a vector or a matrix. We already met other tensors before:

- A **scalar**, or a $(0,0)$ -tensor, is a number for each point in a manifold M . It is of the form of $f(x^\mu)$.
- A **vector**, or a $(1,0)$ -tensor, is a collection of D functions of the coordinates which transforms as:

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

We call the matrix $\frac{\partial x^{\mu'}}{\partial x^\mu}$ the Jacobian matrix. We can write this more familiarly as a Jacobian matrix:

$$J \cdot v,$$

where v is a column vector.

These are the two tensors which we have met so far. We can represent a vector v^μ in a *coordinate-independent way* as a differential operator which takes scalars to tensors:

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

where μ is summed over according to the Einstein summation notation.

- (1) A **one-form** is a set of D functions of the coordinates which transform as:

$$\omega_\mu \rightarrow \omega_{\mu'} = \frac{\partial x^\mu}{\partial x^{\mu'}} \omega_\mu.$$

Note that $\frac{\partial x^\mu}{\partial x^{\mu'}}$ is the *inverse* of the Jacobian. A one form looks similar to a vector. It is known as a **covariant vector**, or a $(0,1)$ -tensor. Note that a one-form has a LOWER index whereas a vector has an UPPER index.

This is useful because if we have a one form ω_μ and a vector v^μ , then we can combine them to make a scalar. To see why, let's **contract** them together:

$$\begin{aligned} v^\mu \omega_\mu &\rightarrow v^{\mu'} \omega_{\mu'} = \left(v^{\mu'} \frac{\partial x^{\mu'}}{\partial x^\mu} \right) \left(\frac{\partial x^\nu}{\partial x^{\nu'}} \omega_\nu \right) \\ &= v^\mu \left(\frac{\partial x^{\mu'}}{\partial x^\mu} \frac{\partial x^\nu}{\partial x^{\nu'}} \right) \omega_\nu \\ &= v^\mu \delta_\mu^\nu \omega_\nu \text{ (by the chain rule)} \\ &= v^\mu \omega_\mu. \end{aligned}$$

This shows that a vector and a one-form get combined together to make a scalar. Formally, we can think of a one-form ω_μ as defining a map from vectors to numbers. We see one-forms referred to as **dual vectors** which live in a **dual tangent space**, $T_p^*(M)$. You've encountered one-forms before. In linear algebra, we usually represent vectors as column vectors and one-forms as row vectors.

$$\omega_\mu = [\omega_1 \quad \cdots \quad \omega_d] \quad v^\mu = \begin{bmatrix} v^1 \\ \vdots \\ v^d \end{bmatrix}.$$

The scalar you would have formed from these two objects is the dot product:

$$\omega_\mu v^\mu = [\omega_1 \quad \cdots \quad \omega_d] \begin{bmatrix} v^1 \\ \vdots \\ v^d \end{bmatrix}.$$

The location of an index – namely, if it's an upper index or a lower index – tells us how an object transforms. When we set an upper and a lower index equal to each other and sum over them, we say that we are **contracting the indices**.

$$\omega_\mu v^\mu \rightarrow \omega_\mu v^\mu.$$

Now consider a scalar f . Then, we can define a one-form whose components are just the derivatives of f in a given coordinate system:

$$\frac{\partial f}{\partial x^\mu} = \partial_\mu f.$$

We can obtain the transformation rule for this by using the chain rule:

$$\partial_\mu f \rightarrow \partial_{\mu'} f = \frac{\partial x^\mu}{\partial x^{\mu'}} \partial_\mu f,$$

which is indeed the multiplication rule that defines a one-form. Just as we write vectors in a basis-independent way as $v = v^\mu \partial_\mu$, we can write one-forms in a basis-independent way as $\omega = \omega_\mu dx^\mu$. We can think of the dx^μ as infinitesimal displacements of the coordinate x^μ . They also form a basis of the cotangent space, and they transform as vectors:

$$\begin{aligned} \omega &= \omega_\mu dx^\mu \\ &= \left(\omega_{\mu'} \frac{\partial x^{\mu'}}{\partial x^\mu} \right) \left(\frac{\partial x^\mu}{\partial x^{\nu'}} dx^{\nu'} \right) \\ &= \omega_{\mu'} dx^{\mu'}, \end{aligned}$$

where the final equality follows from the chain rule. Since the notation is consistent, this proves that $\omega_\mu dx^\mu$ is a basis-independent representation of a one-form. There is a special notation for the one-form $\partial_\mu f$:

$$df = \partial_\mu f dx^\mu.$$

If, for example, $f = x^1$, then this defines, by taking derivatives, the one-form whose components are the derivative of that function:

$$\partial_\mu(x^1) = (1, 0, \dots, 0),$$

which is written as dx^1 . Now we are ready to introduce the general definition of a tensor.

Definition 6.1 (Tensor). A (k, l) -**tensor**, $T^{\mu_1 \dots \mu_k}_{\nu_1 \dots \nu_l}$, is a set of D^{k+l} functions which transform under coordinate transformations in a particular way:

$$T^{\mu'_1 \dots \mu'_k}_{\nu'_1 \dots \nu'_l} = \left(\frac{\partial x^{\mu'_1}}{\partial x^{\mu_1}} \right) \dots \left(\frac{\partial x^{\mu'_k}}{\partial x^{\mu_k}} \right) \left(\frac{\partial x^{\nu_1}}{\partial x^{\nu'_1}} \right) \dots \left(\frac{\partial x^{\nu_l}}{\partial x^{\nu'_l}} \right) T^{\mu_1 \dots \mu_k}_{\nu_1 \dots \nu_l}$$

An easy way to remember Definition 6.1 is to remember that this is the only way to “balance the indices” in a way that respects the Einstein summation convention: upper and lower indices are summed over. We call (k, l) the **rank of the tensor**, and k is the number of **contravariant indices** and l is the number of **covariant indices**. Crucial note: the order and placement of the indices in Definition 6.1 matter!

Sometimes you’ll see people writing tensors in the coordinate-invariant way by contracting it with a bunch of partial derivatives and a bunch of one-forms that are multiplied together by using some cross product as a circle – this is called a **tensor product**, but we’ll leave that formal stuff to the mathematicians to deal with.

$$T = T^{\mu_1 \dots \mu_k}_{\nu_1 \dots \nu_l} \partial_{\mu_1} \otimes \dots \otimes \partial_{\mu_k} \otimes dx^{\nu_1} \otimes \dots \otimes dx^{\nu_l},$$

e.g., we can write a $(0, 2)$ -tensor with components $g_{\mu\nu}$ as:

$$\begin{aligned} g &= g_{\mu\nu} dx^\mu \otimes dx^\nu \\ &= g_{\mu\nu} dx^\mu dx^\nu. \end{aligned}$$

Now, a few comments about tensors. There are three operations that we can do on tensors:

- (1) **Add** together two tensors of the same rank to get another tensor of the same rank:

$$T^{\mu_1 \dots \mu_k}_{\nu_1 \dots \nu_l} + T^{\mu_1 \dots \mu_k}_{\nu_1 \dots \nu_l} = T^{\mu_1 \dots \mu_k}_{\nu_1 \dots \nu_l}.$$

- (2) **Multiply** a tensor with rank (k, l) with a tensor of rank (k', l') to get a tensor of rank $(k + k', l + l')$:

$$T^{\mu_1 \dots \mu_k}_{\nu_1 \dots \nu_l} S^{\mu_{k+1} \dots \mu_{k+k'}}_{\nu_{l+1} \dots \nu_{l+l'}} \text{ is a tensor of rank } (k + k', l + l').$$

- (3) Given a tensor of rank (k, l) , we can **contract** an upper and a lower index together to get a tensor of rank $(k - 1, l - 1)$:

$$T^{\mu_1 \dots \alpha \dots \mu_{k-1}}_{\nu_1 \dots \alpha \dots \nu_{l-1}}.$$

e.g., given a vector and a one-form, you can define a $(1, 1)$ -tensor $v^\mu \omega_\nu$ with μ and ν separate indices, and $v^\mu \omega_\mu$ as a scalar (a $(0, 0)$ -tensor).