Special Relativity

Emre Özer
Winter 2019

Preface.

Contents

Chapter 1

Basics

($c = 1$)

1.1 Postulates

Definition 1. A *reference frame* is a system of coordinates (labels) which associates a position $\vec{x}$ and a time t for *every point in spacetime.*

Definition 2. A frame of reference in which a free body moves with constant velocity is said to be *inertial.*

It follows that if two reference frames move uniformly relative to each other, and if one of them is inertial then so is the other one. Now, we state two experimental facts which we call the postulates of relativity.

Postulate 1. *The laws of physics are the same in all inertial frames.*

Postulate 2. *The speed of light in vacuum is the same in all inertial frames.*

1.2 Interval

Definition 3. An *event* is a point in space time.

We now express the second postulate in a mathematical form. Consider two events $\mathscr{P}_1$ and $\mathscr{P}_2$, connected by a light beam. Let $\mathcal{S}$ be an inertial frame, in which the events have coordinates

$$\mathscr{P}_1 = (t_1, \vec{x}_1), \quad \mathscr{P}_2 = (t_2, \vec{x}_2).$$

Since the two events are connected by a light beam, we have

$$|\vec{x}_2 - \vec{x}_1|^2 = (t_2 - t_1)^2.$$

We write this in the form

$$(t_2 - t_1)^2 - (x_2 - x_1)^2 - (y_2 - y_1)^2 - (z_2 - z_1)^2 = 0.$$

Let $\mathcal{S}'$ be another inertial frame. By postulate two, we immediately have

$$(t'_2 - t'_1)^2 - (x'_2 - x'_1)^2 - (y'_2 - y'_1)^2 - (z'_2 - z'_1)^2 = 0.$$

This motivates us to define an *interval.*

Definition 4. Given an inertial frame $\mathcal{S}$ and two events with coordinates $(t_1, \vec{x}_1), (t_2, \vec{x}_2)$ the interval between them is defined as

$$s_{12} = \left[(t_2 - t_1)^2 - (\vec{x}_2 - \vec{x}_1)^2\right]^{\frac{1}{2}}. \tag{1.1}$$

Note that, by the second postulate, if in any given inertial frame the interval is zero, it must be zero in *every inertial frame*:

$$s = 0 \Longleftrightarrow s' = 0. \tag{1.2}$$

1.2.1 Invariance of the interval

We will now prove that the interval between *any two events* is invariant between inertial frames. But first, we need to show that inertial frames must be related to each other by *linear transformations*. From now on, we denote coordinates by the convention

$$(t, x, y, z) \longrightarrow (x^0, x^1, x^2, x^3) \longrightarrow x^\mu$$

with $\mu = 0, 1, 2, 3$. Similarly for frame $\mathcal{S}'$,

$$(t', x', y', z') \longrightarrow (x^{0'}, x^{1'}, x^{2'}, x^{3'}) \longrightarrow x^{\mu'}.$$

Proposition. The coordinate transformations from an inertial frame to another are *linear*.

Proof. Let $\mathcal{S}$ and $\mathcal{S}'$ be two inertial frames. Consider an arbitrary clock, reading time τ, moving at uniformly in frame $\mathcal{S}$. By homogeneity, equal ticks in τ correspond equal intervals in coordinates $(t, \vec{x})$. Therefore,

$$\frac{dx^\mu}{d\tau} = \text{constant}, \quad \frac{d^2x^\mu}{d\tau^2} = 0.$$

In general, the coordinates of the clock in $\mathcal{S}'$ is given by some function of the coordinates x^μ:

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

By chain rule, we have

$$\frac{dx^{\mu'}}{d\tau} = \frac{dx^\mu}{d\tau}\frac{\partial x^{\mu'}}{\partial x^\mu}$$

and similarly,

$$\begin{aligned} \frac{d^2x^{\mu'}}{d\tau^2} &= \frac{d}{d\tau}\left[\frac{dx^\mu}{d\tau}\frac{\partial x^{\mu'}}{\partial x^\mu}\right] \\ &= \underbrace{\frac{d^2x^\mu}{d\tau^2}}_{=0}\frac{\partial x^{\mu'}}{\partial x^\mu} + \frac{dx^\mu}{d\tau}\frac{dx^\nu}{d\tau}\frac{\partial^2 x^{\mu'}}{\partial x^\mu \partial x^\nu} \\ &= \frac{dx^\mu}{d\tau}\frac{dx^\nu}{d\tau}\frac{\partial^2 x^{\mu'}}{\partial x^\mu \partial x^\nu} = 0. \end{aligned}$$

Since this must hold for all inertial frames $\mathcal{S}'$, we must have

$$\frac{\partial^2 x^{\mu'}}{\partial x^\mu \partial x^\nu} \equiv 0,$$

and so the coordinate transformation $x^{\mu'} = x^{\mu'}(x)$ is linear. □

Proposition. Given any two inertial frames $\mathcal{S}$ and $\mathcal{S}'$, the interval between any two events in $\mathcal{S}$ is equal to the interval in $\mathcal{S}'$.

Proof. Our starting point will be to show that the infinitesimal interval

$$ds^2 = dt^2 - dx^2 - dy^2 - dz^2$$

is of the same order in any two inertial frames. Let's state this precisely. Consider two arbitrary inertial frames $\mathcal{S}$ and $\mathcal{S}'$. Suppose we parameterize the interval between two arbitrary points by some parameter ϵ (this can be done by parameterizing one of the points), such that

$$\lim_{\epsilon\to 0} s(\epsilon) = 0 \Longleftrightarrow \lim_{\epsilon \to 0} s'(\epsilon) = 0. \tag{$*$}$$

As $\epsilon \to 0$, $s(\epsilon) \to ds$ and $s'(\epsilon) \to ds'$ assuming the two points approach each other in the limit $\epsilon \to 0$. Now, the infinitesimal intervals ds and ds' are of the same order if

$$\lim_{\epsilon \to 0} \frac{s(\epsilon)}{s'(\epsilon)} = A \neq 0, \tag{$**$}$$

meaning they approach zero in same order in ϵ. Let's show explicitly that this is the case.

As we are dealing with the interval between two points, we can fix one of the points without loss of generality. So, let's define one of the points to be the origin of $\mathcal{S}$ and $\mathcal{S}'$, denoted $\mathscr{O}$. Now, choose any point in spacetime $\mathscr{P}$ and consider some parameterization $\mathscr{P}(\epsilon)$ such that

$$\lim_{\epsilon \to 0} \mathscr{P}(\epsilon) = \mathscr{O}.$$

Clearly, with such parameterization, we satisfy $(*)$. In frame $\mathcal{S}$, the point $\mathscr{P}(\epsilon)$ will have some coordinates

$$\mathscr{P}(\epsilon) = (t(\epsilon), \vec{x}(\epsilon)),$$

where we consider parameterizations such that the functions $t(\epsilon)$ and $\vec{x}(\epsilon)$ are analytic in ϵ, which we are allowed to do in a general sense because spacetime does not have gaps in it. Then, near $\epsilon = 0$, in general we will have

$$t(\epsilon) = \mathcal{O}(\epsilon^n), \quad x(\epsilon) = \mathcal{O}(\epsilon^m), \quad y(\epsilon) = \mathcal{O}\big(\epsilon^k\big), \quad z(\epsilon) = \mathcal{O}\big(\epsilon^\ell\big),$$

for some $n, m, k, \ell \geq 0$. It then follows that

$$s(\epsilon) = \mathcal{O}\Big(\epsilon^{\min(n,m,k,\ell)}\Big).$$

We note that if one of the indices, (say k), equals zero, then the condition $\mathscr{P}(\epsilon \to 0) = \mathscr{O}$ can only be satisfied if the corresponding coordinate (say y), is identically zero. In this case, it does not contribute to the behaviour of $s(\epsilon \to 0)$. So, when we write $\min(n, m, k, \ell)$ we only consider the non-zero powers.

Now, since the transformation from $x^\mu \to x^{\mu'}$ is linear, we can write it as

$$J^{\mu'}{}_{\mu} x^\mu = x^{\mu'}.$$

We know that we are able to invert this coordinate transformation since all inertial frames are treated on equal footing, hence the determinant of the Jacobian is non-zero. It then follows that there exists at least one coordinate $x^{\mu'}$ such that

$$x^{\mu'} = \mathcal{O}\Big(\epsilon^{\min(n,m,k,\ell)}\Big),$$

and there doesn't exist any $x^{\nu'}$ such that

$$x^{\nu'} = \mathcal{O}(\epsilon^p) \quad \text{with} \quad p < \min(n, m, k, \ell).$$

This is all we need to state that

$$s'(\epsilon) = \mathcal{O}\Big(\epsilon^{\min(n,m,k,\ell)}\Big) = \mathcal{O}(s(\epsilon)).$$

Therefore, for an arbitrary parameterization which is analytic in the coordinates, $(**)$ is satisfied.

Now, let's consider a particular parameterization: $\mathscr{P}(\epsilon) = (\epsilon t_0, \epsilon \vec{x}_0)$ in $\mathcal{S}$. The interval between $\mathscr{P}(\epsilon)$ and $\mathscr{O}$ in frame $\mathcal{S}$ is

$$s^2(\epsilon) = (\epsilon t_0)^2 - (\epsilon \vec{x}_0)^2 \quad \Rightarrow \quad \lim_{\epsilon \to 0} s^2(\epsilon) = dt^2 - dx^2 - dy^2 - dz^2 = ds^2,$$

where as $\epsilon \to 0$, $\epsilon x^\mu \to dx^\mu$. From $(**)$ we have

$$ds^2 = A ds'^2.$$

We don't know what the coefficient A might be, but since it relates two inertial frames the only parameters it can depend on are the coordinates $x^\mu, x^{\mu'}$ and the relative velocity $\vec{v}$ of the frames. By *homogeneity* of space and time, we immediately conclude that there cannot be any dependence on the coordinates - *there are no special points in spacetime.* Also, by *isotropy* of space it cannot depend on the direction of $\vec{v}$ - *space has no preferred direction.* Hence we conclude $A = A(v)$ can only be a function of the magnitude of the relative velocity between frames $\mathcal{S}$ and $\mathcal{S}'$. Note that by choosing a particular parameterization (path) for $\mathscr{P}(\epsilon)$, we essentially convert any dependence of A on the path parameterized by ϵ to the coordinates $(t_0, \vec{x}_0)$, so don't have to worry about A depending on the path we take.

Let v be the speed of $\mathcal{S}$ relative to $\mathcal{S}'$. Then,

$$ds^2 = A(v) ds'^2.$$

But there is nothing special about frame $\mathcal{S}$, and since A only depends on the relative speed, by symmetry we must also have

$$ds'^2 = A(v) ds^2.$$

Together, these imply $A(v) \equiv \pm 1$. Considering a third frame moving relative to $\mathcal{S}$ and $\mathcal{S}'$ it becomes clear that $A(v) \equiv 1$, hence

$$ds^2 = ds'^2. \tag{1.3}$$

Since the infinitesimal intervals are invariant, clearly finite intervals must remain invariant. □

1.2.2 Time-like and space-like separations

Two events are said to be *timelike* separated if their interval is real, meaning $\Delta s^2 > 0$. This immediately implies that there exists an inertial frame in which the two events occur at the same position, setting $\Delta x = 0$,

$$\Delta s^2 = \Delta t^2 - \Delta x^2 = \Delta t^2 > 0.$$

Similarly, two events are said to be *spacelike* if their interval is imaginary, meaning $\Delta s^2 < 0$. Again, it follows that there exists an inertial frame in which the two events happen simultaneously, setting $\Delta t = 0$,

$$\Delta s^2 = \Delta t^2 - \Delta x^2 = -\Delta x^2 < 0.$$

Since the interval is invariant, a timelike interval remains timelike in all inertial frames. Similarly, a spacelike interval remains spacelike. This is

1.3 Proper Time

Imagine a particle, following an arbitrary path $x^\mu(\lambda)$ through spacetime. How is the time measured by the particle related to the time an inertial observer measures?

At any given instant, the particle can be regarded as an inertial frame, in which case it travels a distance $|d\vec{x}|$ in time dt in some reference frame $\mathcal{S}$. Let $d\tau$ be the time experienced by the particle. In the frame of the particle, the displacement is zero since the particle sees itself at rest, so we have

$$ds^2 = dt^2 - dx^2 - dy^2 - dz^2 = d\tau^2,$$

from which we obtain

$$d\tau = dt\sqrt{1 - \frac{d\vec{x}^2}{dt}} = dt\sqrt{1 - v^2},$$

where v is the velocity of the particle in frame $\mathcal{S}$. If we want to obtain the time experienced by the particle over some path, we simply integrate this

$$\tau = \int_\lambda d\tau = \int_{t(0)}^{t(\lambda)} dt\sqrt{1 - v^2(t)}.$$

The time experienced by the moving particle is called the *proper time* of the particle. It equals the interval ds in natural units. Furthermore, it is always the less than the time measured by a different observer. Moving clocks tick slower. As a consequence, since the proper time equals the interval, we obtain that the maximum value for

$$\int_a^b ds$$

is obtained if it is taken along the straight world line joining the two points together.

1.4 Lorentz Transformations

We can obtain transformations which take us from one inertial frame $\mathcal{S}$ with coordinates x^μ to another $\mathcal{S}'$ with coordinates $x^{\mu'}$ by simply considering the invariant interval. This mathematically translates to finding the Jacobian matrix $J^{\mu'}{}_{\mu}$. Landau motivates the form of the group of transformations that leaves the interval invariant (Lorentz group) as follows:

> *"...we may say that the required transformation must leave unchanged all distances in the x, y, z, t space. But such transformations consist only of parallel displacements, and rotations of the coordinate system. Of these the displacement of the coordinate system parallel to itself is of no interest, since it leads only to a shift in the origin of the space coordinates and a change in the time reference point. Thus the required transformation must be expressible mathematically as a rotation of the four-dimensional x, y, z, t coordinate system.*
>
> *"Every rotation in the four-dimensional space can be resolved into six rotations, in the planes xy, zy, xz, tx, ty, tz. The first three of these rotations transform only the space coordinates; they correspond to the usual space rotations."*

So, we expect six transformations in our group, three of which we already now. Let's consider a rotation in the tx plane, which corresponds to a boost in the $x-$direction.

Assuming $\mathcal{S}$ and $\mathcal{S}'$ share their origins, we have a general linear transformation:

$$\begin{aligned} t' &= At + Bx, \\ x' &= Ct + Dx. \end{aligned} \tag{1.4}$$

By the invariance of the interval, we must have

$$s^2 = t'^2 - x'^2 = t^2\big(A^2 - C^2\big) - x^2\big(D^2 - B^2\big) + 2xt(AB - CD) = t^2 - x^2.$$

$$\Rightarrow \quad A^2 - C^2 = 1, \quad D^2 - B^2 = 1, \quad AB = CD.$$

The most general solution is given by

$$A = \pm\cosh\psi, \quad B = \pm\sinh\psi, \quad C = \pm\sinh\psi, \quad D = \pm\cosh\psi$$

for some $\psi \in \mathbb{R}$. If the two frames are identical, we require $x = x'$ and $t = t'$. This immediately gets rid of two minus signs:

$$A = \cosh\psi, \quad D = \cosh\psi.$$

As for the plus and minus ambiguity in B and C, we note that since ψ is arbitrary, we can absorb the sign into ψ. For now, let's take $B = C = -\sinh\psi$ and solve for ψ. So far, we have

$$\begin{aligned} t' &= t\cosh\psi - x\sinh\psi, \quad & t &= t'\cosh\psi + x'\sinh\psi, \\ x' &= x\cosh\psi - t\sinh\psi, \quad & x &= x'\cosh\psi + t'\sinh\psi. \end{aligned} \tag{$*$}$$

Let $\mathcal{S}'$ move with velocity v in $+x$ direction in $\mathcal{S}$. As both frames coincide at the origin, we have

$$x' = 0 \Longleftrightarrow x = vt.$$

Substituting this into $(*)$, we obtain

$$\begin{aligned} t &= t'\cosh\psi \\ vt &= t'\sinh\psi \end{aligned} \quad \Rightarrow \quad v = \tanh\psi.$$

By hyperbolic identities, we have

$$\cosh^2\psi - \sinh^2\psi = 1 \quad \Rightarrow \quad 1 - \tanh^2\psi = \frac{1}{\cosh^2\psi} \quad \Rightarrow \quad \cosh\psi = \sqrt{\frac{1}{1 - v^2}},$$

and similarly we have

$$\sinh\psi = \tanh\psi\cosh\psi = v\sqrt{\frac{1}{1 - v^2}}.$$

Hence, we obtain the Lorentz transformation for a rotation in the tx plane

$$\begin{aligned} t' &= \gamma(v)(t - vx), \quad & t &= \gamma(v)(t' + vx'), \\ x' &= \gamma(v)(x - vt), \quad & x &= \gamma(v)(x' + vt'), \\ dy' &= dy & dy &= dy', \\ dz' &= dz & dz &= dz', \end{aligned} \tag{1.5}$$

where we defined $\gamma(v) = (1 - v^2)^{-\frac{1}{2}}$.

Note that $\gamma(v \to 1)$ diverges. This sets the speed of light as a natural speed limit. Finally, note that in general, Lorentz transformations *do not commute.* As they are rotations in the four dimensional space, the order in which two rotations are performed matters (unless the axis of rotation remains the same).

1.4.1 Length contraction

Suppose there is a rod in frame $\mathcal{S}$, parallel to the $x-$axis. Let its length be $\ell = x_2 - x_1$. We want to determine the rod's length in $\mathcal{S}'$, so we need to measure x_2' and x_1' at the same time t'. We have

$$x_1 = \gamma(x_1' + vt'), \quad x_2 = \gamma(x_2' + vt') \quad \Rightarrow \quad \ell = \gamma\ell'.$$

Since $\gamma > 1$, the length ℓ' in the moving frame is contracted. This is *Lorentz contraction.*

1.4.2 Addition of velocities

Suppose $\mathcal{S}'$ moves with velocity v relative to $\mathcal{S}$ along the x direction. Let $\vec{u}$ be the velocity of a particle in the $\mathcal{S}$ and $\vec{u}'$ be its velocity in $\mathcal{S}'$. How are the two related? We have

$$u^i = \frac{dx^i}{dt}, \quad u^{i'} = \frac{dx^{i'}}{dt'}.$$

By (1.5), we have

$$u_x = \frac{dx}{dt} = \frac{dx' + vdt'}{dt' + vdx'} = \frac{u'_x + v}{1 + vu'_x} \tag{1.6}$$

$$u_y = \frac{dy}{dt} = \frac{dy'}{\gamma(dt' + vdx')} = \frac{\sqrt{1-v^2}u'_y}{1 + vu'_x}, \tag{1.7}$$

$$u_z = \frac{dz}{dt} = \frac{dz'}{\gamma(dt' + vdx')} = \frac{\sqrt{1-v^2}u'_z}{1 + vu'_x}, \tag{1.8}$$

Now, suppose that the particle moves on the $x-y$ plane such that we can decompose its velocity into $u_x = u\cos\theta$ and $u_y = u\sin\theta$. Then, by (1.6) and (1.7) we have

$$\tan\theta = \frac{u'\sqrt{1-v^2}\sin\theta'}{u'\cos\theta' + v.} \tag{1.9}$$

This describes the change in the direction of the velocity. Finally, we consider the special case of the deviation of light from one frame to another. This is called *aberration*. In this case, $u = u' = 1$, so $u_x = \cos\theta$ and $u_y = \sin\theta$. From (1.6) and (1.7) we directly obtain

$$\cos\theta = \frac{\cos\theta' + v}{1 + v\cos\theta'}, \quad \sin\theta = \frac{\sin\theta'\sqrt{1-v^2}}{1 + v\cos\theta'}.$$

For small v, these reduce to the classical expression $\theta' - \theta = v\sin\theta'$.

1.4.3 Doppler effect

Suppose we have two inertial frames $\mathcal{S}'$ and $\mathcal{S}$. Let a light source sit at $\vec{x}' = \vec{0}$ emitting light with wavelength (period) λ'. When an observer sat at $\vec{x} = \vec{0}$ observes the light beam, what wavelength λ will he measure? Let's look at different cases.

Longitudinal

We may imagine two signals, separated in time $\Delta t' = \lambda'$ and in space by $\Delta x' = 0$ in $\mathcal{S}'$. Now, we ask: *how far apart in time are these two signals observed at $x = 0$ in $\mathcal{S}$?* Let's break the whole problem down into four events:

- $\mathscr{A}_1(x = 0, t = t_0)$: the first signal is observed, coordinates given in $\mathcal{S}$.
- $\mathscr{A}_2(x = 0, t = t_0 + \lambda)$: the second signal is observed, coordinates given in $\mathcal{S}$.
- $\mathscr{B}'_1(x' = 0, t' = t'_0)$: first signal is emitted, coordinates given in $\mathcal{S}'$.
- $\mathscr{B}'_2(x' = 0, t' = t'_0 + \lambda')$: second signal is emitted, coordinates given in $\mathcal{S}'$.

To simplify our lives, let's fix the origins of our coordinates such that $t_0 = 0$ in $\mathcal{S}$ and $t'_0 = 0$ in $\mathcal{S}'$. Now, let's write the coordinates of $\mathscr{B}_2$ in $\mathcal{S}$:

$$\mathscr{B}_2(x_2, t_2) = (\gamma(\Delta x' + v\Delta t'), \gamma(\Delta t' + v\Delta x')) = (\gamma v\lambda', \gamma\lambda').$$

All we need to do is calculate when a light signal emitted from $\mathscr{B}_2$ reaches the t-axis in $\mathcal{S}$. Now, we have to specify whether the source $x' = 0$ lies in the $x > 0$ or $x < 0$ region, as this determines the orientation of the light beam. The choice we make will not matter in the end, as long as we are consistent. Suppose the source lies on the $x > 0$ half of the xt plane. Then, the signal emitted from $\mathscr{B}_2(x_2, t_2)$ travels along the line

$$(x - x_2) = -(t - t_2),$$

from which we can read off the time it reaches the observer at $x = 0$:

$$t = x_2 + t_2 = \lambda' \gamma(1 + v) = \lambda' \sqrt{\frac{1+v}{1-v}}.$$

Referring back to the point $\mathscr{A}_2$, we see that $t = \lambda$ and so

$$\lambda = \lambda' \sqrt{\frac{1+v}{1-v}}. \tag{1.10}$$

Transverse

This is a simpler case. We imagine the light beam traveling along the y axis, perpendicular to the relative motion of the two frames. Let the source sit at some $Y > 0$ in the $\mathcal{S}$ frame. By the same construction, we write down the events:

- $\mathscr{A}_1(x, y, t) = (0, 0, t_0)$,
- $\mathscr{A}_2(x, y, t) = (0, 0, t_0 + \lambda)$,
- $\mathscr{B}'_1(x', y', t') = (0, 0, 0)$,
- $\mathscr{B}'_2(x', y', t') = (0, 0, \lambda')$.

Now, the first signal takes time to travel a distance Y. So, let's fix our origin $t = 0$ such that we have $\mathscr{B}_1(x, y, t) = (0, Y, 0)$. This implies $t_0 = Y$. As before, we find B_2 in $\mathcal{S}$ coordinates:

$$\mathscr{B}_2(x_2, y_2, t_2) = (\gamma(\Delta x' + v\Delta t'), Y, \gamma(\Delta t' + v\Delta x')) = (\gamma v \lambda', Y, \gamma \lambda').$$

Now, we need to find the time at which a light beam emitted from $\mathscr{B}_2$ reaches the t-axis. We do the approximation: $Y \gg \gamma v \lambda'$. *This approximation essentially means we are just looking at the transverse component.* If the y separation of the source and the observer is large enough, we can simply ignore the distance the source travels in the x direction in one period. So, we solve the equation for $y = 0$:

$$-(y - Y) = (t - t_2) \quad \Rightarrow \quad Y = t_0 + \lambda - t_2 \quad \Rightarrow \quad \lambda = t_2 = \gamma \lambda'.$$

Hence, we obtain the result

$$\lambda = \lambda' \sqrt{\frac{1}{1-v^2}}. \tag{1.11}$$

General

For the general case, we imagine the source at a distance L away from the observer, at an angle θ from the x axis, moving along the x axis. The events are

- $\mathscr{A}_1(x, y, t) = (0, 0, t_0)$,
- $\mathscr{A}_2(x, y, t) = (0, 0, t_0 + \lambda)$,

- $\mathscr{B}_1'(x', y', t') = (0, 0, 0)$,
- $\mathscr{B}_2'(x', y', t') = (0, 0, \lambda')$.

As before, we first calculate event $\mathscr{B}_1$ in $\mathcal{S}$ coordinates. Defining $t_1 = 0$, and noting that the source is a displacement $\vec{x}_1 = (L\cos\theta, L\sin\theta)$ away from the observer, we obtain:

$$\mathscr{B}_1(x_1, y_1, t_1) = (L\cos\theta, L\sin\theta, 0).$$

Now, we can solve for t_0 by noticing that the light has to travel a distance L to reach the observer. Hence, $t_0 = L$ and so $\mathscr{A}_1 = (0, 0, L), \mathscr{A}_2 = (0, 0, L + \lambda)$. Now, we write down $\mathscr{B}_2$ in $\mathcal{S}$:

$$\mathscr{B}_2(x_2, y_2, t_2) = (\gamma v\lambda' + L\cos\theta, L\sin\theta, \gamma\lambda').$$

The light emitted from the source travels with velocity $\vec{u} = (-\cos\theta, -\sin\theta)$, so the time it takes to reach the observer is $x_2\cos\theta + y_2\sin\theta$. Hence, the time at which the light beam reaches the observer is

$$x_2\cos\theta + y_2\sin\theta + t_2 = L\cos^2\theta + \gamma v\lambda'\cos\theta + L\sin^2\theta + \gamma\lambda' = L + \lambda'\gamma(1 + v\cos\theta).$$

But, we know from $\mathscr{A}_2$ that this time equals $L + \lambda$. Hence we obtain the general result

$$\lambda = \lambda'\frac{1 + v\cos\theta}{\sqrt{1 - v^2}}. \tag{1.12}$$

Let's perform a sanity check. When we set $\theta = 0$, we obtain (1.10). This corresponds to longitudinal shift. When we set $\theta = \pi/2$, we obtain (1.11), which corresponds to the transverse shift. Finally, when $v > 0$, the source is moving away from the observer, so $\lambda > \lambda'$ and the light is red-shifted. When $v < 0$, for θ not too large the light will get blue-shifted. The condition for blue-shift is

$$\frac{1 + v\cos\theta}{\sqrt{1 - v^2}} < 1,$$

where remember that $v < 0$.

1.5 Minkowski Space

We have been talking about labeling events in *space-time* by some coordinates x^μ. From the postulates, we motivated an *interval* and showed its invariance. Now, we consider the more fundamental concept of *geometry.* What space are we living in and what are its properties?

Definition 5. Minkowski space is a four-dimensional space with the metric tensor

$$\eta = \begin{pmatrix} +1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}. \tag{1.13}$$

This is called the *Minkowski metric.*

The form of η is not surprising. To see why, simply look at the only invariant we have: the interval! We note that given a metric g_{ij} and a coordinate system x^i, the line element is given by

$$ds^2 = g_{ij}dx^i dx^j.$$

We know that line elements are invariant under coordinate transformations. Since Lorentz transformations are nothing but coordinate transformations of space-time, this motivates us to define as the line element the square of the interval. The Minkowski metric η then follows when we identify the coordinates of our space as $\vec{x} = (t, \vec{r})$.

Lorentz transformations

We know two fundamental properties of Lorentz transformations: they are linear and they leave the interval invariant. Let's mathematise these properties.

Definition 6. Lorentz transformations are *linear transformations*

$$x^\mu \underset{\Lambda}{\longmapsto} \bar{x}^\mu = \Lambda^\mu{}_\nu x^\nu, \tag{1.14}$$

which leave the Minkowski line element invariant:

$$d\bar{s}^2 \equiv \eta_{\sigma\rho} d\bar{x}^\sigma d\bar{x}^\rho = \eta_{\mu\nu} dx^\mu dx^\nu \equiv ds^2. \tag{1.15}$$

We know $d\bar{x}^\sigma = \Lambda^\sigma{}_\mu dx^\mu$, so we have

$$\eta_{\sigma\rho}\Lambda^\sigma{}_\mu \Lambda^\rho{}_\nu = \eta_{\mu\nu} \quad \Leftrightarrow \quad (\Lambda^\intercal)_\mu{}^\sigma \eta_{\sigma\rho}\Lambda^\rho{}_\nu = \eta_{\mu\nu}. \tag{1.16}$$

This is not surprising at all. We would expect the covariant metric tensor $\eta_{\mu\nu}$ to transform as a covariant tensor - so it does. What is surprising, on the other hand, is a subtlety I glanced over. Note that when we wrote the line element in the $\bar{x}^\mu$ frame, we didn't use a metric of the form $\bar{\eta}_{\mu\nu}$. The reason for this is that the metric takes the same form in all inertial frames! This should be understood as a fundamental constraint on the transformations Λ. These relations can be written in matrix form:

$$\bar{\vec{x}} = \Lambda\vec{x}, \quad \Lambda^\intercal \eta \Lambda = \eta. \tag{1.17}$$

Note that $\Lambda^\mu{}_\nu$ is nothing but the Jacobian associated with the transformations $x^\mu \longmapsto \bar{x}^\mu$. Since the transformations are linear, points transform with the Jacobian and (1.14) holds. There are two Lorentz transformations obeying the properties defined above: rotations in space and Lorentz boosts (rotations in space and time). The two can be written as the matrices:

$$\Lambda_{\text{rot}} = \left(\begin{array}{c|ccc} 1 & 0 & 0 & 0 \\ \hline 0 & & & \\ 0 & & R & \\ 0 & & & \end{array}\right), \quad \Lambda_{\text{x boost}} = \left(\begin{array}{cc|cc} \gamma & -\gamma v & 0 & 0 \\ -\gamma v & \gamma & 0 & 0 \\ \hline 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right). \tag{1.18}$$

An immediate corollary of equation (1.17) is that

$$\Lambda^\intercal \eta \Lambda = \eta \quad \Leftrightarrow \quad \eta = \left(\Lambda^{-1}\right)^\intercal \eta \Lambda^{-1}, \tag{1.19}$$

hence if some Λ is a Lorentz transformation, so is its inverse Λ^{-1}. Furthermore, suppose we have two Lorentz transformations Λ_1 and Λ_2. Look at their composition:

$$(\Lambda_1\Lambda_2)^\intercal \eta \Lambda_1\Lambda_2 = \Lambda_2^\intercal \Lambda_1^\intercal \eta \Lambda_1 \Lambda_2 = \Lambda_2^\intercal \eta \Lambda_2 = \eta. \tag{1.20}$$

So, given any two Lorentz transformations, their composition is also a Lorentz transformation. These relations make it clear that the Lorentz transformations obey a group structure. The set of matrices Λ are a representation of the *Lorentz group.*

Infinitesimal Lorentz transformations

We may write an infinitesimal Lorentz transformation in matrix form as

$$\Lambda = (1 + \omega)$$

where 1 is the identity and ω an infinitesimal. Then, from (1.16) it follows that

$$\begin{aligned} \Lambda^\intercal \eta \Lambda &= (1+\omega)^\intercal \eta (1+\omega) \\ &= (1+\omega^\intercal)\eta(1+\omega) \\ &= \eta + \omega^\intercal \eta + \eta\omega + \mathcal{O}\big(\omega^2\big) = \eta \end{aligned}$$

Hence, we obtain that $(\eta\omega) = -(\eta\omega)^\intercal$ and so $(\eta\omega)$ is *antisymmetric.* An immediate corollary is

$$\omega_{\alpha\beta} = \eta_{\alpha\gamma}\omega^\gamma{}_\beta = -\eta\beta\gamma\omega^\gamma{}_\alpha = -\omega_{\beta\alpha}. \tag{1.21}$$

1.5.1 Four-vectors

Now that we have a metric and a group of transformations, we can talk about *tensors.*

Definition 7. A *four-vector* $\vec{A}$ is an object in Minkowski space with *contravariant* components A^μ which, under a Lorentz transformation Λ, transform as

$$\overline{A}^\mu = \Lambda^\mu{}_\nu A^\nu. \tag{1.22}$$

As usual, the metric η defines a scalar product on the space of four-vectors. The norm of a four-vector is invariant and is defined by

$$A^2 = \eta_{\mu\nu} A^\mu A^\nu = \eta_{\mu\nu} \overline{A}^\mu \overline{A}^\nu. \tag{1.23}$$

The scalar product of two four-vectors is again invariant and is given by

$$\vec{A} \cdot \vec{B} = \eta_{\mu\nu} A^\mu B^\nu = \eta_{\mu\nu} \overline{A}^\mu \overline{B}^\nu. \tag{1.24}$$

Definition 8. A (Lorentz) *covector* is an object with components A_μ which are related to their covariant counterparts by

$$A_\mu = \eta_{\mu\nu} A^\nu. \tag{1.25}$$

This is known as *lowering an index.* Similarly, multiplying by the inverse metric η^{-1} yields

$$A^\mu = \left(\eta^{-1}\right)^{\mu\sigma} \eta_{\sigma\nu} A^\nu = \left(\eta^{-1}\right)^{\mu\sigma} A_\sigma \equiv \eta^{\mu\sigma} A_\sigma, \tag{1.26}$$

where we denote the inverse metric by $\eta^{\mu\nu}$. The placement of the indices is enough to tell the difference. Note that the inverse metric, in this case, equals the original metric.

Inverse transformation

Now, we need to clean up a notational mess involving how the inverse Jacobian Λ^{-1} is written. Usually, we would denote the Jacobian associated with the transformation $x^\mu \longrightarrow \bar{x}^\mu$ as

$$J^\mu{}_\nu = \frac{\partial \bar{x}^\mu}{\partial x^\nu}.$$

Similarly, for the inverse transformation $\bar{x}^\mu \longrightarrow x^\mu$ we have

$$\overline{J}^\mu{}_\nu = \frac{\partial x^\mu}{\partial \bar{x}^\nu}.$$

Note that although we used the same letter J to denote both Jacobians, we have distinguished between them by barring the inverse. In special relativity, however, there is a common *and very criminal* way of notating the inverse transformation. The idea is to make use of equation (1.16) in the following way:

$$\eta_{\sigma\rho} \Lambda^\sigma{}_\mu \Lambda^\rho{}_\nu = \eta_{\mu\nu} \quad \Rightarrow \quad \eta^{\alpha\mu} \eta_{\sigma\rho} \Lambda^\sigma{}_\mu \Lambda^\rho{}_\nu = \eta^{\alpha\mu} \eta_{\mu\nu} = \delta^\alpha{}_\nu.$$

Now, we commit the crime of *raising and lowering the indices of the Jacobian*:

$$\eta^{\alpha\mu} \eta_{\sigma\rho} \Lambda^\sigma{}_\mu \equiv \Lambda_\rho{}^\alpha.$$

Jacobians are not tensors and their indices should not be raised and lowered. It just happens to work in this case. Then, we have

$$\Lambda_\rho{}^\alpha \Lambda^\rho{}_\nu = \delta^\alpha{}_\nu.$$

Notice that this is nothing but the statement that ${\Lambda_\rho}^\alpha$ is the inverse Jacobian:

$$\left(\Lambda^{-1}\right)^\alpha{}_\rho \equiv {\Lambda_\rho}^\alpha.$$

All we've done here is to denote the inverse Jacobian as ${\Lambda_\mu}^\nu$, which is just a convention.

Now we're in a position to discuss how covariant components transform:

$$\overline{A}_\mu = \eta_{\mu\nu}\overline{A}^\nu = \eta_{\mu\nu}{\Lambda^\nu}_\alpha A^\alpha = \eta_{\mu\nu}{\Lambda^\nu}_\alpha \eta^{\alpha\beta}A_\beta = {\Lambda_\mu}^\beta A_\beta. \tag{1.27}$$

Hence, covariant component transform with the inverse Jacobian. This extended to any higher rank tensor: *upper indices transform with* Λ, *lower indices transform with* Λ^{-1}:

$$\overline{T}^{\mu_1\ldots\mu_n}_{\nu_1\ldots\nu_m} = {\Lambda^{\mu_1}}_{\alpha_1}\ldots{\Lambda^{\mu_n}}_{\alpha_n}{\Lambda_{\nu_1}}^{\beta_1}\ldots{\Lambda_{\nu_m}}^{\beta_m}T^{\alpha_1\ldots\alpha_n}_{\beta_1\ldots\beta_m}. \tag{1.28}$$

1.5.2 Curves and tangent vectors

Consider a parameter λ that is monotonically increasing along a path in space-time. We may mathematize the path by the map $\lambda \mapsto x^\mu(\lambda)$. The *tangent vector* to the path at point $x(\lambda_0)$ has components

$$x'^\mu(\lambda_0) = \frac{d}{d\lambda}x^\mu(\lambda)\bigg|_{\lambda=\lambda_0}. \tag{1.29}$$

We may write the interval in terms of the tangent vector as

$$ds^2 = \eta_{\mu\nu}dx^\nu dx^\mu = \eta_{\mu\nu}x'^\mu x'^\nu d\lambda^2. \tag{1.30}$$

Since $d\lambda^2 > 0$, we have the following classifications for the tangent vector at some point $\lambda = \lambda_0$:

$$\eta_{\mu\nu}x'^\mu x'^\nu = \begin{cases} > 0 & \text{timelike,} \\ = 0 & \text{lightlike (null),} \\ < 0 & \text{spacelike.} \end{cases} \tag{1.31}$$

Proposition. The classification of the tangent vector is independent of its parameterization.

Proof. Suppose we have a time-like tangent with some parameterization λ. Changing to another parameterization σ, we have

$$0 < \eta_{\mu\nu}\frac{dx^\mu}{d\lambda}\frac{dx^\nu}{d\lambda} = \eta_{\mu\nu}\frac{dx^\mu}{d\sigma}\frac{dx^\nu}{d\sigma}\left(\frac{d\sigma}{d\lambda}\right)^2.$$

Since $\left(\frac{d\sigma}{d\lambda}\right)^2 > 0$, it follows that the tangent is timelike for any parameterization. This generalizes trivially to spacelike and null tangents. □

Definition 9. A curve whose tangent vector is everywhere *timelike* is called a *timelike curve* (and likewise for lightlike and spacelike curves). A curve whose tangent vector is everywhere timelike or null (i.e. non-spacelike) is called a *causal curve*.

A natural Lorentz-invariant parameterization of *timelike* curves is given by the *proper time* τ, so that $x^\mu = x^\mu(\tau)$. We note that this parameterization only makes sense for timelike curves since the condition $d\tau > 0$ guarantees the parameter τ is monotonically increasing along the curve. A corollary for tangents is

$$\eta_{\mu\nu}\frac{dx^\mu}{d\tau}\frac{dx^\nu}{d\tau} \equiv \eta_{\mu\nu}\dot{x}^\nu\dot{x}^\mu = 1, \tag{1.32}$$

where we denote derivatives with respect to τ with dots. Because τ is Lorentz invariant, $\bar{\tau} = \tau$, tangent vectors $\dot{x}^\mu$ of $\tau-$parameterized curves transform as four-vectors:

$$\dot{\bar{x}}^\mu = \frac{d\bar{x}^\mu}{d\bar{\tau}} = \frac{d}{d\tau}({\Lambda^\mu}_\nu x^\nu) = {\Lambda^\mu}_\nu\dot{x}^\nu \tag{1.33}$$

Finally, we note that spacelike curves are parameterized by $-\tau$, since $-d\tau > 0$.

1.5.3 Four-velocity and acceleration

Definition 10. The four-velocity of a massive particle is defined as the tangent vector to the worldline of the particle, parameterised by proper time:

$$u^\mu(\tau) \equiv \dot{x}^\mu(\tau). \tag{1.34}$$

As we've seen, this transforms as a four-vector.

As a corollary of the definition, the components of the four-velocity are not independent. They must obey the relation

$$d\tau^2 = \eta_{\mu\nu}\frac{dx^\mu}{d\tau}\frac{dx^\nu}{d\tau}d\tau^2 = d\tau^2 u_\mu u^\mu \quad \Rightarrow \quad u_\mu u^\mu = 1. \tag{1.35}$$

Definition 11. The four-acceleration is defined similarly to four-velocity

$$a^\mu = \frac{du^\mu}{d\tau} = \frac{d^2x^\mu}{d\tau^2}. \tag{1.36}$$

Differentiating (1.35), we obtain the relation

$$0 = \frac{d}{d\tau}[u^\mu u_\mu] = a^\mu u_\mu + a_\mu u^\mu = 2a^\mu u_\mu. \tag{1.37}$$

So, $\vec{a}$ and $\vec{u}$ are *always "perpendicular"*.

As an aside, we may write the four-velocity and four-acceleration in terms of their classical correspondents

$$\vec{v} = \frac{d\vec{x}}{dt}, \quad \text{and} \quad \vec{\alpha} = \frac{d\vec{v}}{dt}.$$

Recall that $t = \gamma\tau$ (and $t = x^0$), so we have

$$u^\mu = \frac{dx^\mu}{d\tau} = \frac{dx^\mu}{dt}\frac{dt}{d\tau} = \gamma\begin{pmatrix}1\\ \vec{v}\end{pmatrix}. \tag{1.38}$$

Similarly, for acceleration we have

$$a^\mu = \frac{du^\mu}{d\tau} = \frac{dt}{d\tau}\frac{d}{dt}\begin{pmatrix}\gamma\\ \gamma\vec{v}\end{pmatrix} = \gamma\begin{pmatrix}\alpha v\gamma^3\\ \alpha v\gamma^3\vec{v} + \gamma\vec{\alpha}\end{pmatrix} \tag{1.39}$$

where $\alpha = |\vec{\alpha}|$, $v = |\vec{v}|$, and noting that

$$\frac{d\gamma}{dt} = \alpha v\gamma^3.$$

Chapter 2

Tensor Algebra

2.1 Manifolds and coordinates

We are interested in tensors, which are object defined on *differential manifolds*. Basically, an $n-$dimensional manifold is something which *"locally"* looks like $\mathbb{R}^n$. An example is $2-$sphere S^2. It is globally different than $\mathbb{R}^2$ - it is compact - but locally like $\mathbb{R}^2$. As far as we are concerned, an $n-$dimensional manifold $\mathcal{M}$ is a *set of points* such that each point possesses a set of n coordinates $(x^1, x^2, \ldots, x^n)$.

Sometimes, there is no single coordinate system that covers an entire manifold without degeneracy. Coordinate systems that cover a portion of the manifold are called *coordinate patches*. A set of coordinate patches which covers the whole manifold is an *atlas*.

We are interested in how objects that live in the manifold *transform* under coordinate transformations from one coordinate patch to another.

2.2 Curves and surfaces

Given a manifold, we can define curves and surfaces. There are two ways of doing so: *parametrisation* and *constraints*.

A curve is defined parametrically by

$$x^a = x^a(u), \quad (a = 1, 2, \ldots, n). \tag{2.1}$$

There is a notational subtlety involved here already: x^a on the left hand side corresponds to the *coordinate*, whereas $x^a(\cdot)$ on the right hand side corresponds to a *function* $x^a : I \to \mathbb{R}$ where $u \in I \subset \mathbb{R}$. This duality between a *function* and a *coordinate* being denoted identically will persist throughout the chapter, and it is something to always keep in mind.

A surface of $m < n$ dimensions has m degrees of freedom, so it depends on m parameters and is defined, similar to a curve, by

$$x^a = x^a(u_1, u_2, \ldots, u_m), \tag{2.2}$$

where it is implied that $a = 1, 2, \ldots, n$ from context. If $m = n - 1$, the surface is called a *hypersurface*:

$$x^a = x^a(u_1, u_2, \ldots, u_{n-1}). \tag{2.3}$$

In this case, the $n - 1$ parameters can be eliminated from the n equations in (2.3) can be eliminated to give a single equation:

$$f(x^1, x^2, \ldots, x^n) = 0. \tag{2.4}$$

This is the *constraint* form for defining a hypersurface. Similarly, for a surface of m dimensions, we need $n - m$ constraints:

$$\begin{aligned} f_1(x^1, \dots, x^n) &= 0, \\ f_2(x^1, \dots, x^n) &= 0, \\ &\vdots \\ f_{n-m}(x^1, \dots, x^n) &= 0. \end{aligned} \tag{2.5}$$

Example (Sphere in $\mathbb{R}^3$)**.** Consider a sphere of radius a, embedded in $\mathbb{R}^3$. It has two degrees of freedom, so we need two parameters to parametrise it. Choosing Cartesian coordinates $(x^1, x^2, x^3) = (x, y, z)$, (2.2) becomes:

$$x(\theta, \phi) = a \sin\theta \cos\phi, \quad y(\theta, \phi) = a \sin\theta \sin\phi, \quad z(\theta, \phi) = a \cos\theta,$$

where $\theta \in [0, \pi]$ and $\phi \in [0, 2\pi)$. Noting that this is a hypersurface, we may write a single constraint in the form (2.4):

$$x^2 + y^2 + z^2 = a^2 \quad \Rightarrow \quad f(x, y, z) = x^2 + y^2 + z^2 - a^2 = 0.$$

2.3 Coordinate transformations

The essential point of tensor calculus is that when we make a statement about tensors, we want it to hold for *all coordinate systems.* This is sometimes called the *tensorial* property.

Consider a change of coordinates $x^a \mapsto \bar{x}^a$ given by

$$\bar{x}^a = f^a(x^1, \dots, x^n). \tag{2.6}$$

At this stage, we view coordinate transformations *passively* as assigning to a point on the manifold whose old coordinates are $(x^1, \dots, x^n)$, the new coordinates $(\bar{x}^1, \dots, \bar{x}^n)$. We write this in a more compact form:

$$\bar{x}^a = \bar{x}^a(x), \tag{2.7}$$

where $\bar{x}^a$ denote the n functions $f^n(x^1, \dots, x^n)$.

The *Jacobian matrix* associated with the transformation (2.7) is

$$J^a{}_b \equiv \frac{\partial \bar{x}^a}{\partial x^b} = \begin{bmatrix} \frac{\partial \bar{x}^1}{\partial x^1} & \frac{\partial \bar{x}^1}{\partial x^2} & \cdots & \frac{\partial \bar{x}^1}{\partial x^n} \\ \frac{\partial \bar{x}^2}{\partial x^1} & \frac{\partial \bar{x}^2}{\partial x^2} & \cdots & \frac{\partial \bar{x}^2}{\partial x^n} \\ \vdots & & & \vdots \\ \frac{\partial \bar{x}^n}{\partial x^1} & \frac{\partial \bar{x}^n}{\partial x^2} & \cdots & \frac{\partial \bar{x}^n}{\partial x^n} \end{bmatrix}^a{}_b \tag{2.8}$$

where $J^a{}_b$ denotes the transformation matrix associated with $x \mapsto \bar{x}$ and $\bar{J}^a{}_b$ denotes the one with $\bar{x} \mapsto x$. The determinant of J is the *Jacobian* of the transformation $x \mapsto \bar{x}$:

$$J = \left| \frac{\partial \bar{x}^a}{\partial x^b} \right|.$$

We will always assume the determinant of Jacobian matrix is non-zero, in which case we can invert the transformation law (2.7) to obtain

$$x^a = x^a(\bar{x}). \tag{2.9}$$

From the product rule for determinants, it follows that

$$\bar{J} = \left| \frac{\partial x^a}{\partial \bar{x}^b} \right| = \frac{1}{J}. \tag{2.10}$$

From (2.7), we also define the total differential

$$d\bar{x}^a = \frac{\partial \bar{x}^a}{\partial x^1} dx^1 + \cdots + \frac{\partial \bar{x}^a}{\partial x^n} dx^n = \sum_{b=1}^{n} \frac{\partial \bar{x}^a}{\partial x^b} dx^b \equiv \frac{\partial \bar{x}^a}{\partial x^b} dx^b, \tag{2.11}$$

where, from now on, we adopt the Einstein summation convention: *whenever an index is repeated up and down, it is summed over.* The index a is said to be *free* and the index b is said to be *dummy.* In any tensor expression, the free indices should match on both sides.

We define the Kronecker delta δ^a_b as

$$\delta^a_b = \begin{cases} 1 & a = b, \\ 0 & a \neq b. \end{cases} \tag{2.12}$$

It follows that

$$\frac{\partial \bar{x}^a}{\partial \bar{x}^b} = \frac{\partial x^a}{\partial x^b} = \delta^a_b. \tag{2.13}$$

The Kronecker delta is an example of a *"numerical (or constant)" tensor* - meaning it has the same components in all coordinate systems. This will become clear later on. For now, we note an important relation between the transformation matrices:

$$\delta^a_b = \frac{\partial x^a}{\partial x^b} = \frac{\partial x^a}{\partial \bar{x}^c} \frac{\partial \bar{x}^c}{\partial x^b} = \bar{J}^a{}_c J^c{}_b \quad \Rightarrow \quad \bar{\mathbf{J}}\mathbf{J} = \mathbb{1}, \tag{2.14}$$

as we would expect.

Example (Transformation matrices)**.** Consider a coordinate transformation from Cartesian coordinates $(x^a) = (x, y, z)$ to spherical polar coordinates $(\bar{x}^a) = (r, \theta, \phi)$ in $\mathbb{R}^3$. The transformation matrix associated with $x \mapsto \bar{x}$ is

$$\mathbf{J} = \begin{bmatrix} x/r & y/r & z/r \\ x/(r^2 \tan\theta) & y/(r^2 \tan\theta) & z/(r^2 \tan\theta) - 1/(r \sin\theta) \\ -y/(x^2+y^2) & x(x^2+y^2) & 0 \end{bmatrix},$$

and the transformation matrix of $\bar{x} \mapsto x$ is

$$\bar{\mathbf{J}} = \begin{bmatrix} \sin\theta\cos\phi & \sin\theta\sin\phi & \cos\theta \\ r\cos\theta\cos\phi & r\cos\theta\sin\phi & -r\sin\theta \\ -r\sin\theta\sin\phi & r\sin\theta\cos\phi & 0 \end{bmatrix}.$$

2.3.1 Active and passive transformations

Coordinate transformations can be viewed in two ways: active or passive. We need to define both cases properly to avoid confusion later on.

Let's first consider a simple example. In a one dimensional manifold, take some function $f(x) = x^2$. This is the form of the function expressed in x coordinates. Now, consider a simple transformation: $x \mapsto \bar{x} = x + a$, for some constant a. The passive picture tells us f transforms as follows:

$$f(x) = x^2 = (\bar{x} - a)^2 = \bar{f}(\bar{x}).$$

What this says is that, as far as the manifold is concerned, the function f did not transform at all. We simply changed our coordinates. The functional form for f, expressed in the $\bar{x}$ coordinates, changed to $\bar{f}$ to keep the function itself invariant.

In the active picture, instead of changing our coordinates we simply change the objects on the manifold. In this case, the function f transforms. This is done by relabeling $\bar{x} \to x$ in the expression above, in which case we obtain

$$\bar{f}(x) = (x - a)^2.$$

We think of $\bar{f}(x)$ as a *new function*, actively changed with respect to the manifold by the transformation $x \mapsto \bar{x}$.

Here is the way I think about it: imagine your manifold is a table, on which there is a sheet of grid paper. This grid paper represents a particular choice of a coordinate path. Now, we put a pen, which represents an object defined on the manifold, on the grid paper. A passive transformation is moving the grid paper, holding the pen fixed on the table. An active transformation would be moving the pen, holding the grid paper fixed.

2.4 Contravariant tensors

We want to define a geometrical quantity in terms of the transformation properties of its components under a coordinate transformation. Let's start with an example to motivate the definition.

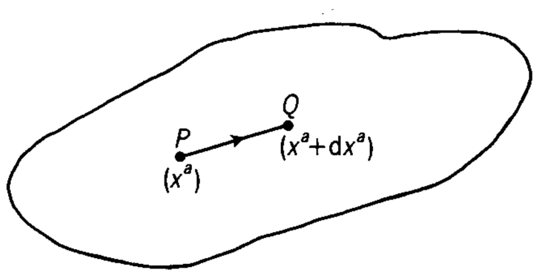


Figure 2.1: Infinitesimal vector at point $\mathcal{P}$.

Consider two points $\mathcal{P}$ and $\mathcal{Q}$ on the manifold, with coordinates (x^a) and $(x^a + dx^a)$ respectively. The two points define an *infinitesimal displacement vector* $\overrightarrow{\mathcal{P}\mathcal{Q}}$. The vector is not *"free"*, it is attached to point $\mathcal{P}$. The components of the vector in $x-$coordinates are dx^a, and in $\bar{x}$-coordinates are $d\bar{x}^a$. The two are related by (2.11):

$$d\bar{x}^a = \frac{\partial \bar{x}^a}{\partial x^b} dx^b. \tag{2.15}$$

The transformation matrix is to be evaluated at point $\mathcal{P}$, so strictly speaking we have

$$d\bar{x}^a = \left.\frac{\partial \bar{x}^a}{\partial x^b}\right|_{\mathcal{P}} dx^b, \tag{2.16}$$

but this will always be implied. This is a linear, homogeneous transformation. With this motivation, we have the following definition:

Definition 12. A *contravariant vector* or *contravariant tensor of rank 1* is a set of n quantities, X^a in (x^a) coordinates, associated with a point $\mathcal{P}$ on the manifold, which transforms under a change of coordinates $x \mapsto \bar{x}$ according to

$$\overline{X}^a = \frac{\partial \bar{x}^a}{\partial x^b} X^b = J^a{}_b X^b, \tag{2.17}$$

where $\partial \bar{x}^a / \partial x^b$ is to be evaluated at $\mathcal{P}$.

An example of a contravariant vector is the infinitesimal displacement dx^a. A contravariant vector with finite components is the *tangent vector* dx^a/du to some curve $x^a(u)$.

Definition 13. A *contravariant tensor of rank 2* is a set of n^2 quantities associated with a point $\mathcal{P}$, denoted T^{ab} in (x^a) coordinates, which transforms according to

$$\overline{T}^{ab} = \frac{\partial \bar{x}^a}{\partial x^c} \frac{\partial \bar{x}^b}{\partial x^d} T^{cd} = J^a{}_c J^b{}_d T^{cd}. \tag{2.18}$$

Higher rank contravariant tensors are defined similarly. An important case is a zero rank tensor, called a *scalar* ϕ, which transforms as

$$\bar{\phi} = \phi \tag{2.19}$$

at $\mathcal{P}$.

As a side note, a *numerical (or constant) tensor* is one which has the same components in every coordinate system. An example is δ^a_b, where this follows by definition. We may check explicitly:

$$\bar{\delta}^a_b = \frac{\partial \bar{x}^a}{\partial x^c} \frac{\partial x^d}{\partial \bar{x}^b} \delta^c_d = \frac{\partial \bar{x}^a}{\partial x^c} \frac{\partial x^c}{\partial \bar{x}^b} = \frac{\partial \bar{x}^a}{\partial \bar{x}^b} = \begin{cases} 1 & a = b, \\ 0 & a \neq b. \end{cases}$$

2.5 Covariant and mixed tensors

We again motivate the definitions by example. Consider a real valued function on the manifold

$$\phi : \mathcal{M} \to \mathbb{R}, \quad x \mapsto \phi(x). \tag{2.20}$$

Note that x stands for the collection $\{x^a\}$. Assuming ϕ is differentiable, consider the coefficients $\partial \phi / \partial x^a$. From (2.7), we have $\phi(x) = \phi(x(\bar{x}))$. By chain rule:

$$\frac{\partial \phi}{\partial \bar{x}^a} = \frac{\partial x^b}{\partial \bar{x}^a} \frac{\partial \phi}{\partial x^b} = \bar{J}^b{}_a \frac{\partial \phi}{\partial x^b}. \tag{2.21}$$

This is the transformation law we wanted to motivate.

Definition 14. A *covariant vector*, or a *covariant tensor of rank 1*, is a set of quantities, denoted X_a in (x^a) coordinates, associated with a point $\mathcal{P}$, which under coordinate transformation $x \mapsto \bar{x}$ transforms according to

$$\overline{X}_a = \frac{\partial x^b}{\partial \bar{x}^a} X_b = \bar{J}^b{}_a X_b. \tag{2.22}$$

Similarly, a covariant tensor of rank 2 is defined by the transformation rule

$$\overline{T}_{ab} = \frac{\partial x^c}{\partial \bar{x}^a} \frac{\partial x^d}{\partial \bar{x}^b} T_{cd} = \bar{J}^c{}_a \bar{J}^d{}_b T_{cd}. \tag{2.23}$$

Contravariant components have *upper*, covariant components have *lower* indices. The fact that coordinate differentials dx^a transform contravariantly is the reason we write the coordinates themselves as x^a rather than x_a. Coordinates, in general, are not tensors.

We can define *mixed tensors* similarly. For example, a mixed tensor of rank 3 with one covariant index and two contravariant indices transforms as

$$\overline{T}_{a}{}^{bc} = \frac{\partial x^i}{\partial \bar{x}^a}\frac{\partial \bar{x}^b}{\partial x^j}\frac{\partial \bar{x}^c}{\partial x^k}T_{i}{}^{jk}. \tag{2.24}$$

If a mixed tensor has contravariant rank p and covariant rank q, it is said to be of *type* (p, q).

Why do we care about tensors and how they transform? Suppose we have a tensor equation in one coordinate system (x^a):

$$R^a{}_b = S^a{}_b.$$

Then, we may multiply both sides by the appropriate transformation matrices to obtain

$$\frac{\partial \bar{x}^a}{\partial x^c}\frac{\partial x^d}{\partial \bar{x}^b}R^c{}_d = \frac{\partial \bar{x}^a}{\partial x^c}\frac{\partial x^d}{\partial \bar{x}^b}S^c{}_d \quad \Leftrightarrow \quad \overline{R}^a{}_b = \overline{S}^a{}_b. \tag{2.25}$$

Hence, if a tensor equation holds in coordinate system, it holds in any coordinate system. This means the mathematical statement we are making with a *tensorial equation* is intrinsic to the manifold and not the coordinate patches. Physics cares only about the manifold, not the coordinates.

2.5.1 Are the transformation matrices tensors?

There is usually confusion, especially in special relativity when we deal with transformation matrices $\Lambda^\mu{}_\nu$, regarding whether these objects are tensors or not. The answer is no. Tensors are geometric objects, intrinsic to the manifold. Transformation matrices tell us how to get from one coordinate patch to another. They are not intrinsic to the manifold.

2.6 Tensor fields

A *tensor field* defined over some region of the manifold is an assignment of a tensor of the same type to every point in the region. It is called *smooth* if its components are differentiable to all orders with respect to the coordinates.

The (passive) transformation law for a contravariant tensor field is

$$\overline{T}^a(\bar{x}) = \left.\frac{\partial \bar{x}^a}{\partial x^b}\right|_{(x)} T^b(x), \tag{2.26}$$

where we note that the transformation matrix is evaluated at (x) - for all points on the manifold where $T^b(x)$ is defined. We will often refer to tensor fields as just tensors and omit the evaluation of the transformation matrix since both will be clear from context.

2.7 Operations with tensors

Proposition. The sum of two tensors of the same type is a tensor of the same type.

Proof. Simply show for a contravariant tensor, generalizes trivially. Let $Z^a = X^a + Y^a$. Then, we have

$$\overline{Z}^a = \overline{X}^a + \overline{Y}^a = \frac{\partial \bar{x}^a}{\partial x^b}X^b + \frac{\partial \bar{x}^a}{\partial x^b}Y^b = \frac{\partial \bar{x}^a}{\partial x^b}\left(X^b + Y^b\right) = \frac{\partial \bar{x}^a}{\partial x^b}Z^a.$$

Since Z^a transforms as a contravariant tensor, result follows. □

A covariant tensor T_{ab} is said to be *symmetric* if $T_{ab} = T_{ba}$. Similar definition holds for contravariant and mixed tensors of rank 2.

Proposition. A symmetric tensor has $\frac{1}{2}n(n+1)$ independent components.

Proof. Thought of as a matrix, the upper half must be equal to the lower half. The number of elements in the upper and lower half is $n^2 - n$ - simply subtracting the diagonal. Hence, the number of independent elements is

$$n^2 - \frac{1}{2}(n^2 - n) = \frac{1}{2}n(n+1).$$

□

The tensor T_{ab} is said to be *antisymmetric* or *skew-symmetric* if $T_{ab} = -T_{ba}$. An immediate corollary is that the diagonal components are identically zero. Hence, an antisymmetric tensor has $\frac{1}{2}n(n-1)$ independent components.

We denote the symmetric part of a tensor by

$$T_{(ab)} \equiv \frac{1}{2}(T_{ab} + T_{ba}), \tag{2.27}$$

and the antisymmetric part by

$$T_{[ab]} \equiv \frac{1}{2}(T_{ab} - T_{ba}). \tag{2.28}$$

In general,

$$T_{(a_1,\ldots,a_m)} \equiv \frac{1}{m!} \times (\text{sum over all permutations of indices}),$$

and

$$T_{[a_1,\ldots,a_m]} \equiv \frac{1}{m!} \times (\text{alternating sum over all permutations of indices}).$$

For example,

$$T_{[abc]} = \frac{1}{6}(T_{abc} - T_{acb} + T_{cab} - T_{bac} + T_{bca} - T_{cba}). \tag{2.29}$$

A *totally symmetric tensor* is a tensor which is equal to its symmetric part. Similarly, a *totally antisymmetric tensor* is one that is equal to its antisymmetric part.

Proposition. A tensor symmetric in a particular coordinate system is symmetric in *any* coordinate system. This means symmetry (or antisymmetry) is an intrinsic property.

Proof. Let $T_{ab} = T_{ba}$. Then,

$$\bar{T}_{ab} = \frac{\partial x^c}{\partial \bar{x}^a}\frac{\partial x^d}{\partial \bar{x}^b}T_{cd} = \frac{\partial x^c}{\partial \bar{x}^a}\frac{\partial x^d}{\partial \bar{x}^b}T_{dc} = \bar{T}_{ba}.$$

This generalizes to any order tensor. □

Proposition. If X^{ab} is antisymmetric and Y_{ab} is symmetric, then $X^{ab}Y_{ab} \equiv 0$.

Proof. Let $\phi = X^{ab}Y_{ab}$. Then

$$\phi = X^{ab}Y_{ab} = -X^{ba}Y_{ba} = -\phi \quad \Rightarrow \quad \phi \equiv 0.$$

□

We can multiply two tensors of type (p_1, q_1) and (p_2, q_2) to obtain a tensor of type $(p_1 + p_2, q_1 + q_2)$. For example,

$$Z^a{}_{bc} = X^a{}_b Y_c. \tag{2.30}$$

A tensor of type (p, q), when multiplied by a scalar field ϕ remains a tensor of type (p, q).

Given a mixed tensor of type (p, q), we can form a tensor of type $(p-1, q-1)$ by *contraction*, which involves summing over an upper and lower index, e.g.

$$X^a{}_{bcd} \longmapsto X^a{}_{acd} = Y_{cd}. \tag{2.31}$$

We can contract a tensor by multiplying it with Kronecker tensor:

$$X^a{}_{acd} = \delta^b_a X^a{}_{bcd}.$$

2.8 Index-free interpretation of contravariant vector fields

We must distinguish between the geometric objects that live in a manifold and their components in particular coordinate systems. Consider a contravariant vector field (similar construction will hold for general tensor fields.) The *key idea* is to treat the vector field as on *operator* which maps real valued fields onto real valued fields. If V is a vector field, for some $f : \mathbb{R}^n \to \mathbb{R}$ we have $Vf = g$ where $g : \mathbb{R}^n \to \mathbb{R}$ is some other field.

In the (x^a) coordinates, we introduce the notation

$$\partial_a \equiv \frac{\partial}{\partial x^a}.$$

Then, V is defined as the operator

$$V = V^a \partial_a. \tag{2.32}$$

Let's look at how V transforms under coordinate transformation $\bar{x} \mapsto x$:

$$\overline{V} = \overline{V}^a \bar{\partial}_a = \frac{\partial \bar{x}^a}{\partial x^b} \frac{\partial x^c}{\partial \bar{x}^a} V^b \partial_c = \epsilon^c_b V^b \partial_c = V^b \partial_b = V. \tag{2.33}$$

Hence, vector field V as an operator is *invariant.* This is why it is geometric.

In any coordinate system, for a given point $\mathcal{P}$ on the manifold, we can think of $[\partial/\partial x^a]_{\mathcal{P}}$ as forming a *basis for all vectors as* $\mathcal{P}$, since any vector at $\mathcal{P}$ is given by (2.32)

$$V_{\mathcal{P}} = [V^a]_{\mathcal{P}} \left[\frac{\partial}{\partial x^a} \right]_{\mathcal{P}}, \tag{2.34}$$

a linear combination of the $[\partial/\partial x^a]_{\mathcal{P}}$. The vector space of all the contravariant vectors at $\mathcal{P}$ is the *tangent space* at $\mathcal{P}$, denoted $T_{\mathcal{P}}(\mathcal{M})$.

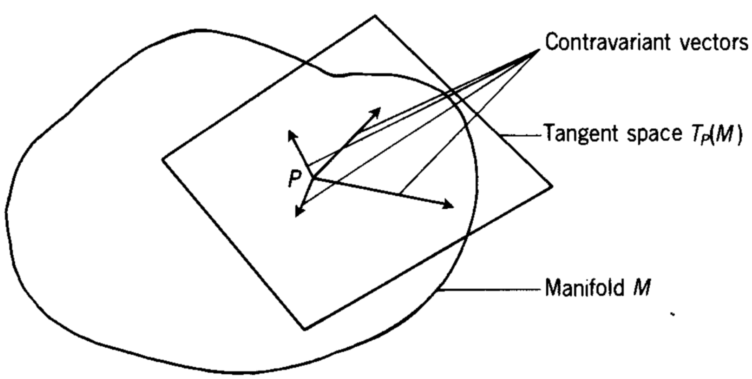


Figure 2.2: The tangent space at $\mathcal{P}$.

In general , the tangent space at any point in a manifold is different from the underlying manifold. As an example, consider the surface of a sphere. The tangent vector at any point does not lie on the sphere. Two exceptions to this are Euclidean and Minkowski spaces.

2.9 Lie brackets

Definition 15. Given two vector fields X and Y, we can define a new vector field called the *commutator* or *Lie bracket* of X and Y by

$$[X, Y] = XY - YX. \tag{2.35}$$

Letting $[X, Y] = Z$ and operating with it on some arbitrary function f, we have

$$\begin{aligned}
Zf &= [X, Y]f \\
&= (XY - YX)f \\
&= X(Y^a\partial_a)f - Y(X^a\partial_a)f \\
&= X^b\partial_b(Y^a\partial_a)f - Y^b\partial_b(X^a\partial_a)f \\
&= (X^b\partial_bY^a)\partial_af + X^bY^a\partial_b\partial_af - (Y^b\partial_bX^a)\partial_af - Y^bX^a\partial_b\partial_af \\
&= (X^b\partial_bY^a - Y^b\partial_bX^a)\partial_af,
\end{aligned}$$

where the $\partial_a\partial_bf$ terms cancel assuming partial derivatives commute. Since f is arbitrary, we obtain the result:

$$Z^a = [X, Y]^a = X^b\partial_bY^a - Y^b\partial_bX^a, \tag{2.36}$$

from which it is clear that Z is a vector field itself.

Corollary. The following results follow from the definition:

1. $[X, X] = X^2 - X^2 \equiv 0.$
2. $[X, Y] = XY - YX = -(YX - XY) \equiv -[Y, X].$
3. Jacobi's identity:

$$\begin{aligned}
[X, [Y, Z]] + [Z, [X, Y]] + [Y, [Z, X]] &= X(YZ - ZY) - (YZ - ZY)X + Z(XY - YX) \\
&\quad - (XY - YX)Z + Y(ZX - XZ) - (ZX - XZ)Y \\
&= XYZ - XZY - YZX + ZYX + ZXY - ZYX \\
&\quad - XYZ + YXZ + YZX - YXZ - ZXY + XZY \\
&\equiv 0.
\end{aligned}$$

Chapter 3

Tensor Calculus

3.1 Partial derivatives of a tensor

We ask which differential operations are *tensorial*? Let's first see how the partial derivative of a tensor, say $\partial_c A^a$, transforms:

$$\bar{\partial}_c \overline{A}^a = \frac{\partial}{\partial \bar{x}^c}\left[\frac{\partial \bar{x}^a}{\partial x^b} A^b\right] = \frac{\partial x^d}{\partial \bar{x}^c}\frac{\partial}{\partial x^d}\left[\frac{\partial \bar{x}^a}{\partial x^b} A^b\right] = \frac{\partial x^d}{\partial \bar{x}^c}\frac{\partial \bar{x}^a}{\partial x^b}\partial_d A^b + A^b \frac{\partial x^d}{\partial \bar{x}^c}\frac{\partial^2 \bar{x}^a}{\partial x^d \partial x^b}. \tag{3.1}$$

What a disaster! This, obviously does not transform like a tensor due to the second term. It is not actually a disaster, in fact there is a fundamental reason why $\partial_c A^a$ is not a tensor in general. By definition, differentiation something involves comparing a quantity evaluated at two neighbouring points. For instance, for a contravariant vector field we compute

$$\lim_{\epsilon \to 0} \frac{[A^a]_{\mathcal{P}} - [A^a]_{\mathcal{Q}}}{\epsilon}$$

where $\mathcal{P}$ and $\mathcal{Q}$ are separated by some distance ϵ. However, from the transformation law we have

$$[\overline{A}^a]_{\mathcal{P}} = \left[\frac{\partial \bar{x}^a}{\partial x^b}\right]_{\mathcal{P}} A^b_{\mathcal{P}}, \quad [\overline{A}^a]_{\mathcal{Q}} = \left[\frac{\partial \bar{x}^a}{\partial x^b}\right]_{\mathcal{Q}} A^b_{\mathcal{Q}}.$$

This involves the transformation matrix evaluated at *different* points. This is the reason derivatives of tensor fields, in general, are not tensors.

An immediate corollary of the discussion above is that if the transformations we consider have transformation matrices that do not depend on the point on the manifold they are evaluated at, then any derivative of a tensor field with respect to any coordinates is a tensor. This is a special case which will become clear later.

We will first introduce the *Lie derivative*, then define *affine connection* and *metric connection* to introduce the *covariant derivative.*

3.2 Lie derivative

Pay attention, this will be involved.

Let's start with a motivation. We wish to 'differentiate' a tensor field along a curve, meaning we want the points $\mathcal{P}$ and $\mathcal{Q}$ to lie on some curve $x^a(u)$. Now, instead of defining a set of curves that cover a portion of the manifold every time we wish to differentiate something, let's first look at how we can refer to $x^a(u)$ by contravariant vector fields.

Consider a *congruence of curves*, defined over the manifold such that only one curve goes through each point. Then, given any one curve $x^a(u)$, we can define the tangent vector field

dx^a/du along the curve. If we do this for every curve in the congruence, we obtain a contravariant vector field $X^a(x)$.

Conversely, given a non-zero vector field $X^a(x)$, we may define a congruence of curves called the *orbits* or *trajectories* of X^a. These curves are obtained by solving

$$\frac{dx^a}{du} = X^a(x(u)). \tag{3.2}$$

This is an ODE, so existence and uniqueness theorems ensure there is a unique solution.

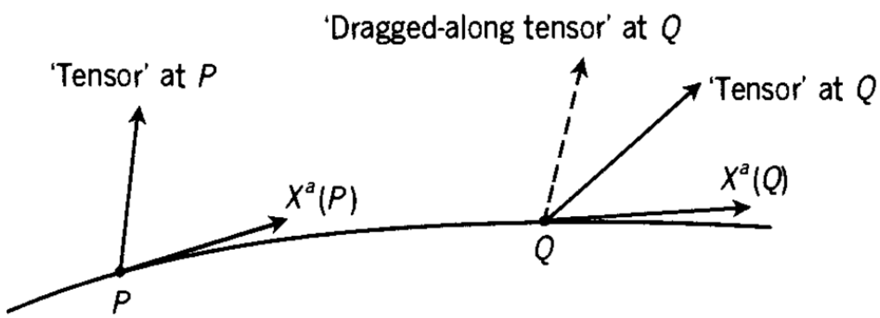
That is the first motivation sorted! Now, suppose we want to differentiate some tensor field $T^{a\cdots}_{b\cdots}(x)$. We already know that if we evaluate this tensor field at different points on the manifold, the resulting difference will not be a tensor. Our second motivation is to fix that.

The essential idea is to use the congruence of curves to *drag the tensor* at some point $\mathcal{P}$ along the curve passing through $\mathcal{P}$ to the neighbouring point $\mathcal{Q}$. Then, we compare this *dragged along* tensor with the original tensor field evaluated at $\mathcal{Q}$. This *dragging* is done by viewing the coordinate transformation from $\mathcal{P}$ to $\mathcal{Q}$ *actively*.

Consider the transformation

$$x^a \longmapsto \bar{x}^a = x^a + \epsilon X^a(x), \tag{3.3}$$

where ϵ is small. This is called a *point transformation*. It sends the point $\mathcal{P}$ with coordinates x^a to the point $\mathcal{Q}$ with coordinates $x^a + \epsilon X^a(x)$. Since the transformation is viewed actively, the coordinates of each point are given in the same coordinate system x^a.



Note that the point $\mathcal{Q}$, by definition, lies on the curve of congruence of X^a through point $\mathcal{P}$. Differentiating (3.3) yields

$$\frac{\partial \bar{x}^a}{\partial x^b} = \delta^a_b + \epsilon \partial_b X^a(x). \tag{3.4}$$

Now, consider some tensor field T^{ab}. Its components at $\mathcal{P}$ are $T^{ab}(x)$ and, under the point transformation (3.3) they transform as

$$\begin{aligned}\overline{T}^{ab}(\bar{x}) = \frac{\partial \bar{x}^a}{\partial x^c}\frac{\partial \bar{x}^b}{\partial x^d}T^{cd}(x) &= (\delta^a_c + \epsilon\partial_c X^a)(\delta^b_d + \epsilon\partial_d X^b)T^{cd} \\ &= T^{ab} + \epsilon[\partial_c X^a + \partial_d X^b]T^{cd} + \mathcal{O}(\epsilon^2),\end{aligned} \tag{3.5}$$

where note that every T^{ab} and X^a on the right hand side is evaluated at point $\mathcal{P}$ with coordinates x.

$\overline{T}^{ab}(\bar{x})$ is the tensor dragged along from $\mathcal{P}$ to $\mathcal{Q}$. We need to compare that with the tensor already at $\mathcal{Q}$, which we express as $T^{ab}(\bar{x})$. This is given by Taylor expanding:

$$T^{ab}(\bar{x}) = T^{ab}(x^c + \epsilon X^c) = T^{ab}(x) + \epsilon X^c \partial_c T^{ab}(x) + \mathcal{O}(\epsilon^2). \tag{3.6}$$

Now, we are in a position to define the Lie derivative of T^{ab} with respect to X^a, denoted $L_X T^{ab}$:

$$L_X T^{ab}(x) = \lim_{\epsilon \to 0} \frac{T^{ab}(\bar{x}) - \overline{T}^{ab}(\bar{x})}{\epsilon}. \tag{3.7}$$

We are comparing the tensor field evaluated at $\mathcal{Q}$ with the tensor field dragged from $\mathcal{P}$ to $\mathcal{Q}$. By (3.5) and (3.6), we have

$$L_X T^{ab} = X^c \partial_c T^{ab} - T^{ac} \partial_c X^a - T^{cb} \partial_c X^a. \tag{3.8}$$

This is a particular example for a contravariant tensor of rank two. For a general tensor field $T^{a\dots}_{b\dots}$, we define the Lie derivative in exactly the same manner:

$$L_X T^{a\dots}_{b\dots}(x) = \lim_{\epsilon \to 0} \frac{T^{a\dots}_{b\dots}(\bar{x}) - \overline{T}^{a\dots}_{b\dots}(\bar{x})}{\epsilon}. \tag{3.9}$$

Since both tensor fields are evaluated at the same point on the manifold, the Lie derivative transforms as a tensor.

3.2.1 Properties of the Lie derivative

For all of the following propositions, it will be easier to rewrite (3.9) as

$$T^{a\dots}_{b\dots}(\bar{x}) - \overline{T}^{a\dots}_{b\dots}(\bar{x}) = \epsilon L_X T^{a\dots}_{b\dots}(\bar{x}) + \mathcal{O}\big(\epsilon^2\big). \tag{3.10}$$

Proposition. It is *linear*. Meaning, for any μ, λ constant we have

$$L_X(\lambda Y^{a\dots}_{b\dots} + \mu Z^{a\dots}_{b\dots}) = \lambda L_X Y^{a\dots}_{b\dots} + \mu L_X Z^{a\dots}_{b\dots}. \tag{3.11}$$

Proof. By (3.10),

$$\begin{aligned} \epsilon L_X(\lambda Y^{a\dots}_{b\dots} + \mu Z^{a\dots}_{b\dots}) + \mathcal{O}\big(\epsilon^2\big) &= \lambda Y^{a\dots}_{b\dots} + \mu Z^{a\dots}_{b\dots} - \lambda \overline{Y}^{a\dots}_{b\dots} - \mu \overline{Z}^{a\dots}_{b\dots} \\ &= \lambda\big(Y^{a\dots}_{b\dots} - \overline{Y}^{a\dots}_{b\dots}\big) + \mu\big(Z^{a\dots}_{b\dots} - \overline{Z}^{a\dots}_{b\dots}\big) \\ &= \epsilon(\lambda L_X Y^{a\dots}_{b\dots} + \mu L_X Z^{a\dots}_{b\dots}) + \mathcal{O}\big(\epsilon^2\big). \end{aligned}$$

Hence, the result follows. □

Proposition. It obeys the Leibniz rule for differentials of products:

$$L_X(Y^{a\dots}_{b\dots} Z^{c\dots}_{d\dots}) = Y^{a\dots}_{b\dots}(L_X Z^{c\dots}_{d\dots}) + Z^{c\dots}_{d\dots}(L_X Y^{a\dots}_{b\dots}). \tag{3.12}$$

Proof. We have

$$Y^{a\dots}_{b\dots} Z^{c\dots}_{d\dots} - \overline{Y}^{a\dots}_{b\dots} \overline{Z}^{c\dots}_{d\dots} = L_X(Y^{a\dots}_{b\dots} Z^{c\dots}_{d\dots})\epsilon + \mathcal{O}\big(\epsilon^2\big), \tag{$*$}$$

and

$$\overline{Y}^{a\dots}_{b\dots} = Y^{a\dots}_{b\dots} - L_X Y^{a\dots}_{b\dots} + \mathcal{O}\big(\epsilon^2\big), \quad \overline{Z}^{c\dots}_{d\dots} = Z^{c\dots}_{d\dots} - L_X Z^{c\dots}_{d\dots} + \mathcal{O}\big(\epsilon^2\big). \tag{$**$}$$

Substitute $(**)$ into $(*)$

$$\begin{aligned} L_X(Y^{a\dots}_{b\dots} Z^{c\dots}_{d\dots})\epsilon + \mathcal{O}\big(\epsilon^2\big) &= Y^{a\dots}_{b\dots} Z^{c\dots}_{d\dots} - (Y^{a\dots}_{b\dots} - L_X Y^{a\dots}_{b\dots})(Z^{c\dots}_{d\dots} - L_X Z^{c\dots}_{d\dots}) + \mathcal{O}\big(\epsilon^2\big) \\ &= \epsilon[Y^{a\dots}_{b\dots}(L_X Z^{c\dots}_{d\dots}) + Z^{c\dots}_{d\dots}(L_X Y^{a\dots}_{b\dots})] + \mathcal{O}\big(\epsilon^2\big). \end{aligned}$$

The result follows. □

Proposition. It is type preserving, meaning the Lie derivative of a tensor of type (p, q) is again a tensor of type (p, q). This is inherited from addition and needs no proof.

Proposition. It commutes with contraction, meaning

$$\delta^b_a L_X T^a{}_b = L_X T^a{}_a\,. \tag{3.13}$$

Proof.

$$\delta^b_a L_X T^a{}_b = \delta^b_a \lim_{\epsilon\to 0} \frac{T^a{}_b - \overline{T}^a{}_b}{\epsilon} = \lim_{\epsilon\to 0} \frac{\delta^b_a (T^a{}_b - \overline{T}^a{}_b)}{\epsilon} = \lim_{\epsilon\to 0} \frac{T^a{}_a - \overline{T}^a{}_a}{\epsilon} = L_X T^a{}_a\,.$$

□

3.2.2 Lie derivatives of various tensor fields

We've seen the Lie derivative of a contravariant tensor of rank two is given by (3.8). Let's now look at other tensor fields and extend our results for a general tensor field.

Scalar field

Let ϕ be a scalar field. Then, it transforms as

$$\overline{\phi}(\bar{x}) = \phi(x),$$

and by Taylor expanding:

$$\phi(\bar{x}) = \phi(x^a + \epsilon X^a) = \phi(x) + \epsilon X^a \partial_a \phi + \mathcal{O}\big(\epsilon^2\big).$$

By equation (3.7) we have

$$L_X \phi = \lim_{\epsilon\to 0} \frac{\phi(\bar{x}) - \overline{\phi}(\bar{x})}{\epsilon} = \lim_{\epsilon\to 0} \frac{\phi(x) + \epsilon X^a \partial_a \phi - \phi(x) + \mathcal{O}\big(\epsilon^2\big)}{\epsilon} = X^a \partial_a \phi. \tag{3.14}$$

Contravariant vector field

Let T^a be a contravariant vector field. The transformation law is

$$\overline{T}^a(\bar{x}) = \frac{\partial \bar{x}^a}{\partial x^b} T^b(x) = (\delta^a_b + \epsilon \partial_b X^a) T^b = T^a + \epsilon T^b \partial_b X^a,$$

and by Taylor expanding

$$T^a(\bar{x}) = T^a(x^b + \epsilon X^b) = T^a(x) + \epsilon X^b \partial_b T^a + \mathcal{O}\big(\epsilon^2\big).$$

Hence, by (3.7)

$$L_X T^a = \lim_{\epsilon\to 0} \frac{T^a(\bar{x}) - \overline{T}^a(\bar{x})}{\epsilon} = X^b \partial_b T^a - T^b \partial_b X^a = [X, T]^a. \tag{3.15}$$

Conveniently, the Lie derivative of a contravariant vector field T with respect to the contravariant vector field X is their Lie bracket $[X, T]$.

Covariant vector field

First, we rewrite the transformation law (3.3):

$$x^a = \bar{x}^a + \epsilon X^a,$$

from which it follows that

$$\frac{\partial x^b}{\partial \bar{x}^a} = \delta^b_a - \epsilon \overline{\partial}_a X^b = \delta^b_a - \epsilon \frac{\partial x^c}{\partial \bar{x}^a} \partial_c X^b = \delta^b_a - \epsilon(\delta^c_a - \epsilon X^c)\partial_c X^b = \delta^b_a - \epsilon \partial_a X^b + \mathcal{O}\big(\epsilon^2\big).$$

Now, let T_a be a covariant vector field. The transformation law is

$$\overline{T}_a(\bar{x}) = \frac{\partial x^b}{\partial \bar{x}^a} T_b(x) = T_a - \epsilon T_b \partial_a X^b + \mathcal{O}\big(\epsilon^2\big).$$

We also have

$$T_a(\bar{x}) = T_a(x^b + \epsilon X^b) = T_a(x) + \epsilon X^b \partial_b T_a + \mathcal{O}\big(\epsilon^2\big).$$

Then, it follows that

$$L_X T_a = \lim_{\epsilon \to 0} \frac{\overline{T}(\bar{x}) - T(\bar{x})}{\epsilon} = X^b \partial_b T_a + T_b \partial_a X^b. \tag{3.16}$$

General tensor field

We generalize (3.15) and (3.16) to write the Lie derivative of a general tensor field $T^{a\dots}_{b\dots}$. First, we note that there will be a term like

$$+X^c \partial_c T^{a\dots}_{b\dots}$$

independent of the type of the tensor field $T^{a\dots}_{b\dots}$. Next, from (3.15) we can see that every contravariant index picks up a term like

$$-T^{c\dots}_{b\dots} \partial_c X^a.$$

Finally, from (3.16) we see that every covariant index picks up a term

$$+T^{a\dots}_{c\dots} \partial_b X^c.$$

Then, we can write the Lie derivative of $T^{a\dots}_{b\dots}$:

$$L_X T^{a\dots}_{b\dots} = X^c \partial_c T^{a\dots}_{b\dots} - T^{c\dots}_{b\dots} \partial_c X^a - \dots + T^{a\dots}_{c\dots} \partial_b X^c + \dots \tag{3.17}$$

3.3 Covariant differentiation and affine connection

We now want to define a limiting process which reduces to ordinary differentiation with respect to Cartesian coordinates in Euclidean space, but still retains its tensorial property. We've already seen something of the form $\partial_b X^a$ does not transform as a tensor, since we evaluate the vector field at different points on the manifold. To get around this, we introduce *parallel transport*.

3.3.1 Parallel transport

Consider a covariant vector field $X_a(x)$. Let points $\mathcal{P}$ and $\mathcal{Q}$ have coordinates x^a and $x^a + dx^a$. We know that

$$X_a(x + dx) - X_a(x) = dX_a = \partial_j X_i \, dx^j \tag{3.18}$$

is not a tensor since the two terms on the left hand side are evaluated at different points. Suppose we somehow "parallel transport" $X_a(x)$ from $\mathcal{P}$ to $\mathcal{Q}$, such that at $\mathcal{Q}$ is has components $X_a + \delta X_a$. Now, we may compare the two vectors defined at the same point:

$$X_a(x + dx) - \left(X_a(x) + \delta X_a\right)\Big|_{\mathcal{Q}} = dX_a - \delta X_a \equiv \nabla_b X_a \, dx^b \,, \tag{3.19}$$

where we defined the *covariant derivative* of a covariant vector field $\nabla_b X_a$. Since the two terms on the left hand side are defined at the same point, it follows that the covariant derivative transforms as a tensor.

Note that all we've done so far was to claim that some "parallel transport" procedure exists such that the components $X_a + \delta X_a$ evaluated at $\mathcal{Q}$ are that of a vector parallel (in some sense) to the one with components X_a evaluated at $\mathcal{P}$. We still need to define what this parallel transport procedure is.

In Euclidean space with Cartesian coordinates, parallel transport should do nothing. This is not surprising, as the ordinary derivative $\partial_a X_a$ transforms as a tensor in such coordinates. Then, suppose our manifold is Euclidean and let (y^a) be Cartesian coordinates. Denote the components of our vector field by Y_a in these coordinates. Then,

$$Y_a = \frac{\partial x^b}{\partial y^a} X_b, \quad X_a = \frac{\partial y^b}{\partial x^a} X_b. \tag{3.20}$$

With parallel displacement, $\delta Y_a = 0$. Hence,

$$\delta X_a = \delta\left(\frac{\partial y^b}{\partial x^a}\delta Y_b\right) = \delta\left(\frac{\partial y^b}{\partial x^a}\right) Y_b = \frac{\partial^2 y^b}{\partial x^a \partial x^c}\, dx^c\, Y_b. \tag{3.21}$$

Substituting back for Y_b:

$$\delta X_a = \frac{\partial^2 y^b}{\partial x^a \partial x^c}\, dx^c\, \frac{\partial x^d}{\partial y^b} X_d = \Gamma^d_{ac}\, dx^c\, X_d, \tag{3.22}$$

where

$$\Gamma^d_{ac} = \frac{\partial^2 y^b}{\partial x^a \partial x^c} \frac{\partial x^d}{\partial y^b} \tag{3.23}$$

in Euclidean space. This motivates us to define δX_a in any manifold as some *bilinear form* of X_a and dx^c with coefficients Γ^a_{bc} so that $\delta X_b = \Gamma^a_{bc} X_a\, dx^c$.

The covariant derivative of a covariant vector field is given by

$$\nabla_b X_a = \partial_b X_a - \Gamma^c_{ab} X_c. \tag{3.24}$$

As a final note, what we're doing with parallel transport is simply taking into account the change in the basis vectors going from point $\mathcal{P}$ to $\mathcal{Q}$. The tangent spaces at these two points are in general different from each other, which is the main reason why we can't compare vectors living at distinct points. By parallel transport, we express the components of the vector at $\mathcal{P}$ in terms of the new basis at $\mathcal{Q}$.

3.3.2 Affine connection

Demanding the covariant derivative transform as a $(0, 2)$ tensor yields the transformation law for Γ^c_{ab}:

$$\bar{\Gamma}^c_{ab} = \frac{\partial x^i}{\partial \bar{x}^a} \frac{\partial x^j}{\partial \bar{x}^b} \frac{\partial \bar{x}^c}{\partial x^k} \Gamma^k_{ij} + \frac{\partial^2 x^k}{\partial \bar{x}^a \partial \bar{x}^b} \frac{\partial \bar{x}^c}{\partial x^k}. \tag{3.25}$$

Any object Γ^c_{ab} with such transformation property is called an *affine connection*. It is easy to show that an equivalent transformation law is

$$\bar{\Gamma}^c_{ab} = \frac{\partial x^i}{\partial \bar{x}^a} \frac{\partial x^j}{\partial \bar{x}^b} \frac{\partial \bar{x}^c}{\partial x^k} \Gamma^k_{ij} - \frac{\partial x^k}{\partial \bar{x}^b} \frac{\partial x^\ell}{\partial \bar{x}^a} \frac{\partial^2 \bar{x}^c}{\partial x^\ell \partial x^k}. \tag{3.26}$$

3.3.3 Covariant derivatives of tensors

Consider a scalar field ϕ. It is obviously unaffected by parallel transport, so $\delta\phi = 0$. Hence, we have

$$\nabla_a \phi = \partial_a \phi, \tag{3.27}$$

which we already know transforms as a type $(0,1)$ tensor.

Now, consider a contravariant vector field A^a. We can construct a scalar by contraction $B_a A^a$. Then, we know

$$0 = \delta(A^a B_a) = A^a \delta B_a + B_a \delta A^a \quad \Rightarrow \quad B_a \delta A^a = -A^a \Gamma^c_{ab} B_c \, dx^b \,. \tag{3.28}$$

Since B_a is arbitrary, it follows that

$$\delta A^a = -\Gamma^a_{bc} A^b \, dx^c \,. \tag{3.29}$$

Hence,

$$\nabla_c A^a \, dx^c = dA^a - \delta A^a = \partial_c A^a + \Gamma^a_{bc} A^b \, dx^c \,. \tag{3.30}$$

This defines the covariant derivative for a contravariant vector field:

$$\nabla_c A^a = \partial_c A^a + \Gamma^a_{bc} A^b. \tag{3.31}$$

Similarly, for a tensor field $T^a{}_b$, consider the parallel displacement of the scalar $T^a{}_b A^b B_a$, from which it follows that

$$\nabla_c T^a{}_b = \partial_c T^a{}_b + \Gamma^a_{dc} T^d{}_b - \Gamma^d_{bc} T^a{}_d \,. \tag{3.32}$$

Covariant differentiation is obviously linear. We can check that is it Leibniz. Let $C^a = A^a_b B^b$:

$$\begin{aligned}
\nabla_c C^a &= \partial_c C^a + \Gamma^a_{bc} C^b \\
&= \partial_c \big(A^a_b B^b\big) + \Gamma^a_{bc} A^b_d B^d \\
&= (\partial_c A^a_b) B^b + A^a_b \partial_c B^b + \Gamma^a_{bc} A^b_d B^d \\
&= A^a_b \big(\partial_c B^b + \Gamma^b_{dc} B^d\big) - A^a_b \Gamma^b_{dc} B^d + \Gamma^a_{bc} A^b_d B^d + (\partial_c A^a_b) B^b \\
&= A^a_b \big(\nabla_c B^b\big) + B^b \big(\partial_c A^a_b - \Gamma^d_{bc} A^a_d + \Gamma^a_{dc} A^d_b\big) \\
&= A^a_b \big(\nabla_c B^b\big) + (\nabla_c A^a_b) B^b.
\end{aligned} \tag{3.33}$$

Finally, note that the difference of two affine connections is a tensor since the inhomogeneous terms cancel. The anti-symmetric part of Γ^a_{bc} is called the *torsion tensor*:

$$T^a_{bc} = \Gamma^a_{bc} - \Gamma^a_{cb}. \tag{3.34}$$

If the torsion tensor vanishes, the connection is symmetric in its lower indices.