General Relativity with insights

Daniel Lin

April 19, 2023

This note on General Relativity will provide useful insights into understanding general relativity, though it may not pursue strict mathematical proofs. To read this note, you should be familiar with vectors, linear transformations, and matrices.

Due to the large amount of work required to prepare these notes, and the difficulty of explaining the very abstract concept of general relativity, I did not finish writing these notes. But still, I hope these semi-finished notes will help you through the first few weeks of studying the course Tensor calculus and General relativity.

Contents

1 Special Relativity

1.1 Moving frame and Galilean Transformation

The theory of relativity is always concerned with the space-time frame, i.e. $(x.y, z, t)$. For any dimension of space, time will be the last index. Later, you will see why this is useful. Let's focus on one-dimensional space-time frame (x, t). Suppose you stand at the origin and observe a moving object passing at a constant speed of 1 unit per second. We could draw the trajectory(or a decent name, the *world line*) of this object on space-time frame as shown in figure 1. (Be careful that since we are drawing (x, t) instead of (t, x), objects with higher speed trace out a trajectory with a lower gradient, i.e. closer to the x-axis)

Of course, you would view yourself as stationary, so your trajectory will be the vertical line $x \equiv 0$. We are interested in how the moving object would observe you. From now on, [frame 0] means your perspective, and [frame 1] means the perspective of the moving object. To find [frame 1], we must transform the moving object's trajectory to the line $x \equiv 0$.

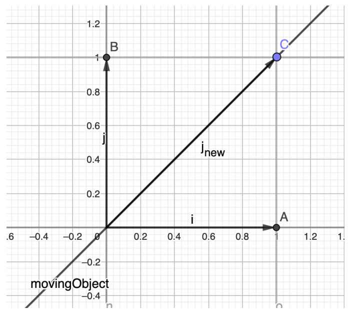


Figure 1: Observing a moving object with speed 1 [frame 0]

There are many ways to do this transformation, but we need to state a few requirements that feel like common sense:

- Straight lines should remain straight lines. So one object moving at a constant speed in one frame(i.e. the trajectory is a straight line) will not accelerate in another. This ensures Newton's first law is not violated
- Relative speed will not change. i.e. in your view, the object is moving at speed 1, so in the object's view, you should be moving backwards at speed 1. i.e. our trajectory in [frame 1] should be a straight line with gradient -1.
- Time-universal: the clock of all objects run simultaneously. Note that point B in figure 1 means your position at time $t = 1$, and point C means the object's position at time $t = 1$. In your view, if your clock is at $t = 1$, the object's clock should also be pointing to $t = 1$. i.e. point B and C should be on the same horizontal level.

These requirements leave us with few choices. We need a linear transformation that will not move the horizontal lines but move the object's trajectory to t-axis. If you already have an idea, you can check figure 2. This is a shear mapping along the x direction, and the transformation matrix is

$$\begin{pmatrix} x' \\ t' \end{pmatrix} = \begin{pmatrix} 1 & -v \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ t \end{pmatrix}$$

where $v = 1$ is the speed of the moving object.

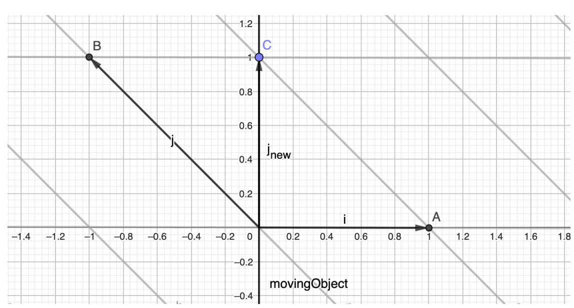


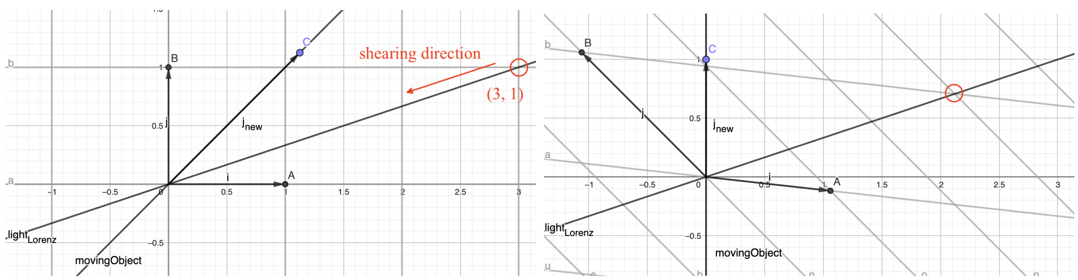
Figure 2: Galilean transformation to the object's view [frame 1]

Now your trajectory is along the line crossing O, B. And indeed, the gradient is -1. All requirements are satisfied. This new frame is moving with the moving object. However, Michelson–Morley experiment shows that the speed of light c is constant regardless of relative motion. This shear changes the speed of every moving object (on the space-time diagram, this means the gradient changed), including light. If we transform the trajectory $(cs, s), s \in \mathbb{R}$ of a beam of light moving forward in [frame 0]:

$$\begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} cs \\ s \end{pmatrix} = \begin{pmatrix} cs - s \\ s \end{pmatrix}$$

the speed in [frame 1] becomes $c-1$, the speed of light is not constant.

Suppose you drop the time-universal assumption(i.e. the horizontal lines can be no more horizontal) and design a transformation that preserves the speed of light. In that case, you get the *Lorentz transformation*. Geometrically, this is easy to achieve, shear along the trajectory of light instead of x-axis. See figure 3. (The figure is drawn assuming the speed of light $c=3$)



(a) Before Lorentz transform [frame 0] **(b) After Lorentz transform [frame 1]**

Figure 3: Geometric demonstration of Lorentz transformation

We will derive the algebraic expression of Lorentz transformation later because more details are required to satisfy the assumption "relative speed does not change". Once the correct parameters are chosen, as shown in figure 3b, the trajectory of the moving object is transformed to x-axis, and the angle between your trajectory and the object's trajectory remains unchanged.

Before you throw away Galilean transformation, please look at figure 4 where $c=20, v=0.4$.

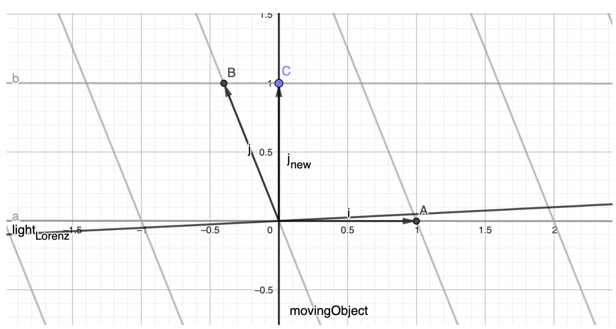


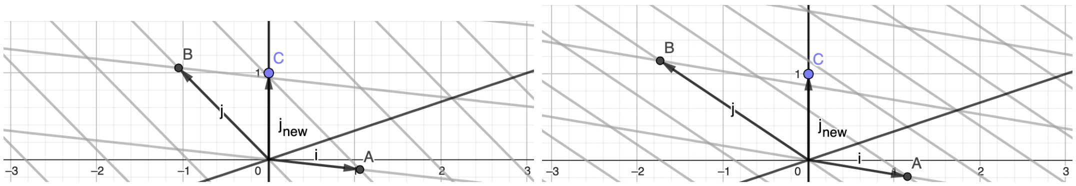
Figure 4: Lorentz transformation with $c=20, v=0.4$

The transformation looks just like the Galilean transformation! Because the speed of light is high compared to the moving object, so the trajectory of light is close to the x-axis. This means the shear can almost be viewed as a shear along x-axis. So for the motion of objects with a much slower speed than light, there is no need to worry about the theory of relativity. But in figure 3, where the object's speed is $1/3$ of that of light, things become more interesting.

I created a Geogebra demonstration of Lorentz transformation which can be found on `https://www.geogebra.org/classic/ynuz6kxf`. There are few adjustable parameters: c is the speed of light, $v \in [-c, c]$ is the speed of frame and $s \in [-c, c]$ will control the speed of moving object. Lorentz transformation only uses the speed v. If you let $v = s$, you will be in [frame 1].

1.2 Properties of Lorentz transformation

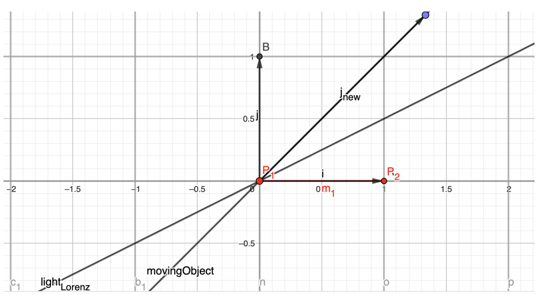
Firstly, time is not universal! Observe points B and C, which means the position of you and the moving object after the time elapsed for 1 unit, respectively; they are at different heights! In the object's view, when the object's clock points to $t = 1$, your clock is not $t = 1$ yet, maybe $t = 0.9$. (geometrically, this means point B is higher than point C) But in [frame 0], your clock is faster than the moving object's! Because point C is higher than point B. This is called time dilation. Although you and the object both feel the time is faster than the other, there is no contradiction. As you can see geometrically, this is just the effect of shear mapping. Faster objects have more time dilation, as shown in figure 5.



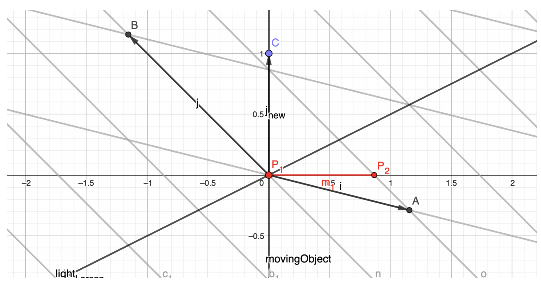
(a) $c = 3, v = 1$ **(b)** $c = 3, v = 1.5$

Figure 5: Lorentz transformation for two different object speed

Another effect brought by Lorentz transformation is length contraction. Suppose an object has two endpoints P_1, P_2 with length 1 unit, and it is stationary in [frame 0]. See the length after transformation in figure 6, it becomes moving backwards in [frame 1] so the length becomes shorter. The fact that a moving object becomes shorter is called *Length Contraction*.



(a) Before Lorentz transformation [Frame 0]



(b) After Lorentz transformation [Frame 1]

Figure 6: Lorentz transformation and length contraction

1.3 The mathematical deductions

Based on the requirement that straight lines remain straight lines, we consider the linear transformation with parameters $\alpha, \beta, \gamma, \delta$

$$\begin{pmatrix} x' \\ t' \end{pmatrix} = \underbrace{\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}}_{=:T_L} \begin{pmatrix} x \\ t \end{pmatrix}$$

We need to move the trajectory $x = vt$ to $x' = 0$, and move the trajectory $x = 0$ to $x' = -vt'$:

$$\begin{pmatrix} 0 \\ t' \end{pmatrix} = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \begin{pmatrix} vt \\ t \end{pmatrix}, \quad \begin{pmatrix} -vt' \\ t' \end{pmatrix} = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \begin{pmatrix} 0 \\ t \end{pmatrix}$$

Solving equations yields

$$\beta = -\alpha v, \quad \beta = -\delta v \quad \Rightarrow \delta = \alpha$$

so the transformation becomes

$$\begin{pmatrix} x' \\ t' \end{pmatrix} = \begin{pmatrix} \alpha & -\alpha v \\ \gamma & \alpha \end{pmatrix} \begin{pmatrix} x \\ t \end{pmatrix}$$

Now using the assumption that the speed of light is constant, we plug in $x = ct$, $x' = ct'$, and will obtain

$$\gamma = -\frac{v}{c^2}\alpha, \quad T_L = \alpha \begin{pmatrix} 1 & -v \\ -\frac{v}{c^2} & 1 \end{pmatrix}$$

We have used all the assumptions, but yet one parameter α to solve. Here, we use the fact that Lorentz transformation should be invertible, first using matrix inverse:

$$\begin{pmatrix} x \\ t \end{pmatrix} = \frac{1}{\alpha D} \begin{pmatrix} 1 & \frac{v}{c^2} \\ v & 1 \end{pmatrix} \begin{pmatrix} x' \\ t' \end{pmatrix}$$

$D := 1 - v^2/c^2 = \det T_L$. $D \neq 0$ for the matrix to have an inverse, so $v \in [-c, c]$. i.e. the speed of any object cannot exceed that of light. But we know there is another way to find the inverse transformation. Because transforming from [frame 1] to [frame 0] should be another Lorentz transformation but with speed $-v$ instead,

$$\begin{pmatrix} x \\ t \end{pmatrix} = \alpha \begin{pmatrix} 1 & v \\ v/c^2 & 1 \end{pmatrix} \begin{pmatrix} x' \\ t' \end{pmatrix}$$

Two inverses must align, so $\frac{1}{\alpha D} = \alpha \Rightarrow \alpha = 1/\sqrt{D} = 1/\sqrt{1 - v^2/c^2}$.

When $v \ll c$, $\alpha \approx 1$, $-v/c^2 \approx 0$, so

$$T_L \approx 1 \begin{pmatrix} 1 & -v \\ 0 & 1 \end{pmatrix}$$

which is the Galilean transformation.

Exercise 1. *Suppose two events take place at the same point x, but at different time t_1, t_2. i.e. the two events are $(x, t_1), (x, t_2)$ in the space-time. Find the difference in time after Lorentz transformation w.r.t $\Delta t := t_2 - t_1$ (it should be independent of x) Do you see why time dilation occurs?*

Now suppose two events actually take place at different points x_1, x_2, what is the effect of $\Delta x := x_2 - x_1$ on the difference in time after Lorentz transformation?

Use similar techniques, and study why Lorentz brings length contraction. (i.e. find a relation between moving length $L' := \Delta x'$ and the static length $L := \Delta x$)

Summary

The Lorentz transformation by components is given by

$$x' = \gamma(x - vt), \quad t' = \gamma(-\frac{v}{c^2}x + t) \tag{1}$$

Inverse Lorentz:

$$x = \gamma(x' + vt'), \quad t = \gamma(\frac{v}{c^2}x' + t') \tag{2}$$

$$t' = \sqrt{1 - \frac{v^2}{c^2}}t \tag{3}$$

Time dilation (observed time of a moving object from a static observer will be $1 - v^2/c^2$ times)
If the observed length of an object by a static (relative to the object) observer is L, then the observed length by a moving observer of speed v will be

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

1.4 Twins paradox

Suppose twins A and B (of the same age) both live on earth. One day, B suddenly wants to travel to star S. He travels at speed v to star S and then suddenly changes direction, coming back to earth at speed $-v$. Suppose before B departs, they each calibrate an accurate clock C_A, C_B respectively. In A's view, B is moving at a high speed, so the clock C_B should be slower than C_A. (i.e. when B comes back, B should be younger than A) But in B's view (i.e. in an inertia frame moving with B), A travels at a high speed, so C_A is slower than C_B. A contradiction! If B never goes back to earth, these two claims can both hold because A and B are standing at different points. But at the moment A reunites with B, the clock's comparison should have no ambiguity. Actually, it turns out that A's view is correct.

The knowledge we learnt about special relativity does not apply directly to B because a frame moving with B is NOT an inertia frame. B has a sudden acceleration at star S. Equivalently, I am saying that the view of B's frame in figure 7 is wrong.

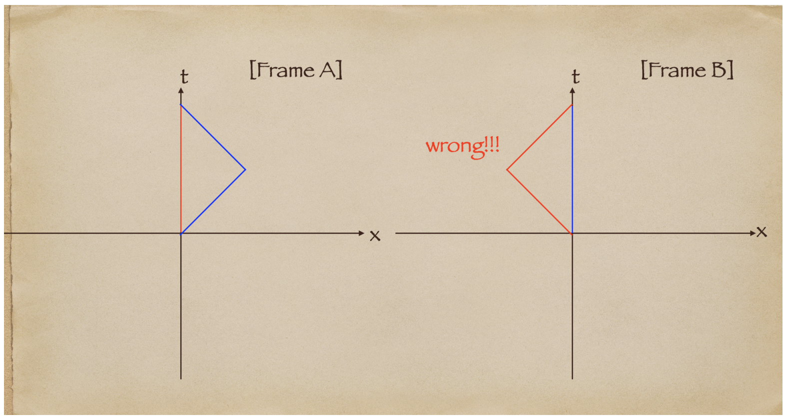


Figure 7: A wrong perspective of the twin's paradox

Note in this section, the red line will be the world line/trajectory of A, and the blue line will be the world line of B.

There are two reasons why figure 7 is wrong. Firstly, this is not a Lorentz transformation (from [frame A] to [frame B]). Secondly, even if no force acts on A (or the Earth), in [frame B], it suddenly accelerates. This breaks Newton's first law. But [frame A] is indeed correct and we will keep on using it.

The world line of B should really be divided into two parts: the journey from Earth to S, and from S to Earth. (And the instantaneous turn, which we will discuss later) During both journeys, B is moving at a constant speed, so we can use two inertia frames [frame B1] and [frame B2] to represent them respectively.

Assume E (the earth) and S are stationary to each other. Consider the journey from the Earth to the star. If in A's view, the distance between E and S is l, then by length contraction, B would view the distance to be rl. (where the ratio $r := \sqrt{1 - v^2/c^2}$). So the observed A's time by A is l/v, and observed B's time by B is rl/v. How do they observe each other's time? We need a Lorentz transformation (see figure 8) Note I added a new green line representing the world line of star S. The dotted lines sl_B, sl_A are simultaneous lines for B and A (i.e. sl_B are lines $t' = k$, sl_A are lines $t = k$ where k is a constant) By equation 1, suppose $t' = 0$, then $t = v/c^2 x$. Similarly, you can deduce that other simultaneous lines of B in [frame A] also have a gradient v/c^2. By symmetry, sl_A in [frame B1] have gradient $-v/c^2$ (you can deduce this rigorously by plugging $t = k$ to equation 2)

Suppose now we put a clock C_S on start S. When B arrives at S, the clock on C_S should point to l/v, because C_S, C_A are relatively stationary. But as you can see from the highest sl_B in [frame A], observed A's elapsed time by B during B's journey is slightly shorter than l/v, it should be

$$\frac{l}{v} - \frac{v}{c^2} l$$

because gradient of sl_B are all v/c^2. By time dilation (equation 3, observed B's travel time by A is rl/v. This is the same as observed B's travel time by B, although the ratio r comes from length contraction for B. The following table summarises the time elapsed for A and B during the journey to star S.

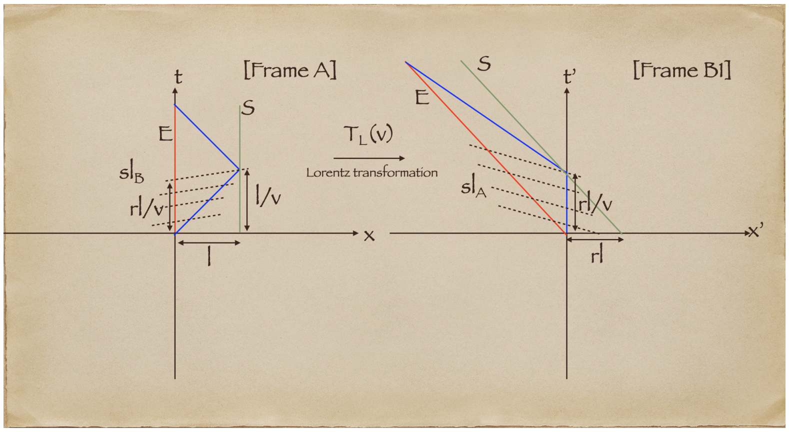


Figure 8: Lorentz transformation for the journey from E to S

	time elapsed for A	time elapsed for B
observed by A	l/v	rl/v
observed by B	$\frac{l}{v} - \frac{v}{c^2}l$	rl/v

note $l/v - vl/c^2 = l/v(1 - v^2/c^2) = r^2 l/v < rl/v$. So actually A and B both think the other's time is slower. (This is possible because they are not at the same point in space-time)

Now the key point to solve the twin's paradox is that, as B turns around, changing speed from v to $-v$, the gradient of sl_B suddenly changes from v/c^2 to $-v/c^2$, as shown in figure 9. It is not hard to find that t-intercept of the two red lines (representing sl_B before and after the turning) differs by $2\frac{v}{c^2}l$.

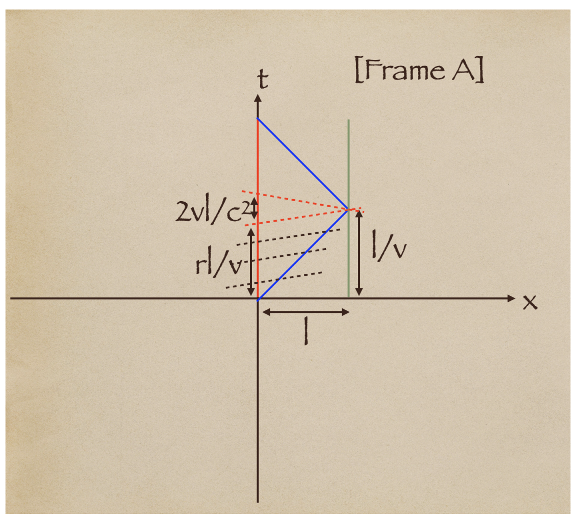


Figure 9: Sudden change of sl_B in [Frame A]

The following table summarises the time elapsed for A and B when B turns at star S.

	time elapsed for A	time elapsed for B
observed by A	0	0
observed by B	$2\frac{v}{c^2}l$	0

Note in A's view, this turn is also instantaneous. So time elapsed for A observed by A will be 0.

This difference is still present even if the turn is gradual. Because ultimately, the gradient of sl_B has to change from v/c^2 to $-v/c^2$ in order to enter [frame B2]. Let's draw [frame B2], by putting the point at which B arrives at star S at the origin for [frame B1] and apply Lorentz transformation. (see figure 10) I kept one simultaneous line in [Frame B1] (and marked it in red) and applied Lorentz transformation to it as well. In [Frame B2], you can see how sharply the simultaneous line suddenly changes.

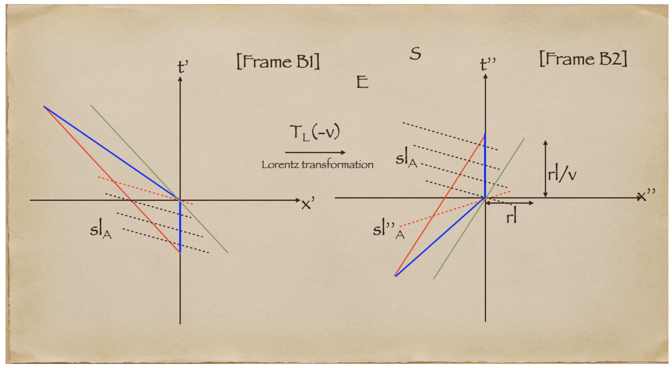


Figure 10: Lorentz transformation from Frame B1 to B2

Finally, since the journey from S back to Earth is an exact copy of the journey from Earth to S (simply with speed reversed), so the time elapsed will not be different. The following table summarises the time elapsed for A and B during the journey from S to Earth.

	time elapsed for A	time elapsed for B
observed by A	l/v	rl/v
observed by B	$\frac{l}{v} - \frac{v}{c^2}l$	rl/v

Finally, we can add all numbers in three tables and obtain the desired result: the following table summarises the time elapsed for A and B for the whole journey

	time elapsed for A	time elapsed for B
observed by A	$2\frac{l}{v}$	$2r\frac{l}{v}$
observed by B	$2\frac{l}{v} - 2\frac{v}{c^2}l + 2\frac{v}{c^2}l = 2\frac{l}{v}$	$2r\frac{l}{v}$

There is no paradox at all! The ratio of time elapsed by B and time elapsed by A is $r = \sqrt{1 - v^2/c^2}$, so B will be younger than A.

The twins' paradox actually applies to any journey path. In general, if A moves from event P_1 to P_2 in a straight line, and B's journey does not take the shortest path, you can always first apply a Lorentz transformation so that the world line of A is along the t axis, and then draw simultaneous lines sl_B on [frame A]. Figure 11 is an example. This phenomenon is called *Principle of Maximal Ageing*.

1.5 Index Notation, Lorentz group

The general Lorentz transformation can be written in index notation. It is STRONGLY suggested that you read the second chapter Introduction to Tensors before you continue reading this chapter.

Matrices are (1, 1) tensors, and they are capable of dealing with one vector and one co-vector. So we write components of Lorentz transformation matrix as $\Lambda^{\mu'}_{\nu}$ and using Einsteins' notation, we can write $x^{\mu'} = \Lambda^{\mu'}_{\nu} x^{\nu}$ where x^{ν} $(\nu = 1, 2, 3)$ are components of the original frame defined by

$$x^0 = ct,\ x^1 = x,\ x^2 = y,\ x^3 = z$$

this is usually called the *Minkowski space*. Scaling by c on x^0 ensures that the world line of a photon would be on the cone

$$(x^0)^2 = (x^1)^2 + (x^2)^2 + (x^3)^2$$

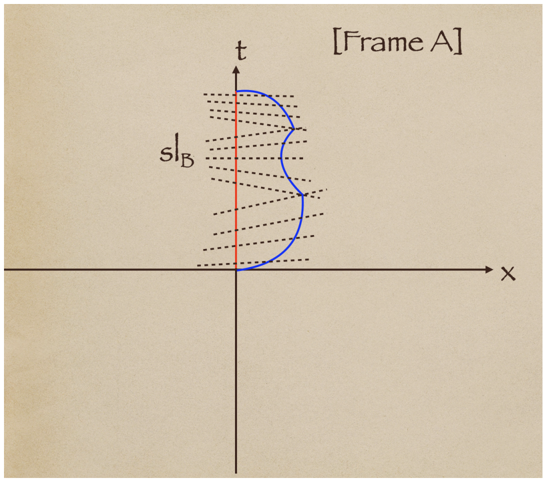


Figure 11: An example of the principle of maximal ageing

where the 2 outside brackets are squaring, whereas numbers on top of x are indices. This cone is called *Null Cone*. The space enclosed by the cone contains all the possible future world lines of a particle standing at the origin. (see figure 12)

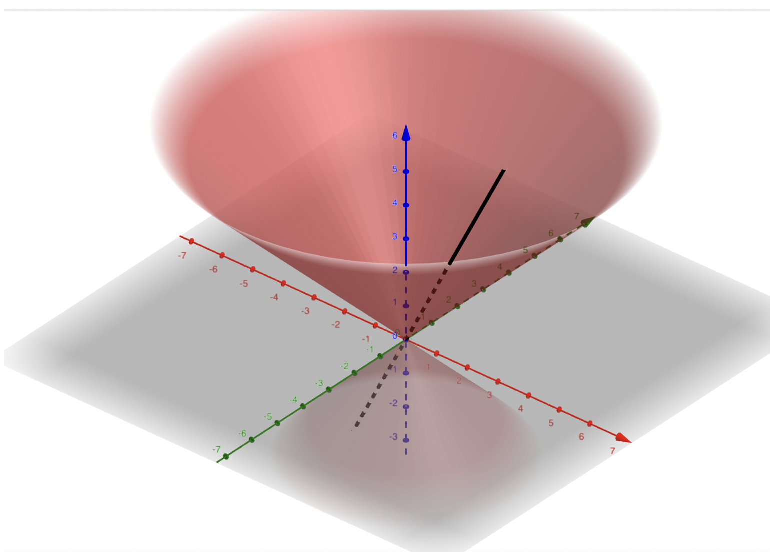


Figure 12: Null cone and a possible world line

Vectors inside the null cone are *time-like*, vectors outside the null cone are *space-like* and vectors along the cone are *null*. The tangent vector(sometimes called *world velocity*) of the world line of an object is either time-like or null. For photons, the world velocity is always null.

If you are given the vector component, you can distinguish between them using the inner product

$$v \cdot v \begin{cases} > 0, & \text{time-like} \\ = 0, & \text{null or } v = 0 \\ < 0, & \text{space-like} \end{cases}$$

Note the inner product here is defined using the metric η, i.e. $v \cdot v = v^T \eta v$.

Another advantage of these new coordinates is that now the matrix for Lorentz transformation discussed before is symmetric:

$$\begin{pmatrix} x' \\ t' \end{pmatrix} = \gamma \begin{pmatrix} 1 & -v \\ -v/c^2 & 1 \end{pmatrix} \begin{pmatrix} x \\ t \end{pmatrix}, \quad \begin{pmatrix} x^{0'} \\ x^{1'} \end{pmatrix} = \gamma \underbrace{\begin{pmatrix} 1 & -v/c \\ -v/c & 1 \end{pmatrix}}_{=:\Lambda} \begin{pmatrix} x^0 \\ x^1 \end{pmatrix}$$

x^2, x^3 are omitted because they are not changed. In this case, Lorentz transformation can be represented using rapidity ψ (defined s.t. $\tanh\psi = v/c$)

$$\Lambda = \begin{pmatrix} \cosh\psi & -\sinh\psi \\ -\sinh\psi & \cosh\psi \end{pmatrix}$$

this looks like a rotation, and actually, rotations are members of the Lorentz group (will be defined later)

Recall the differential time-dilation formulae (or *proper time formulae*)

$$\tau = \sqrt{(1 - \frac{v^2}{c^2}}t$$

here τ is called *proper time*. The equivalent form is (as derived in official lecture notes in equation (5))

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

. And in the new coordinate system, this is

$$ds^2 = (dx^0)^2 - (dx^1)^2 - (dx^2)^2 - (dx^3)^2 \tag{4}$$

where $s := c\tau$. LHS is a quadratic form of vector dx and it can be written in matrix form (which is a special case of the bilinear form). But we will adopt the tensor notation, writing equation 4 as

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

where $\eta_{\mu,\nu}$ are components of metric tensor(a quadratic form) η. There are two indices on the bottom because it is capable of dealing with two vectors. It is not hard to find that

$$\eta_{ij} = \begin{cases} 1, & i = j = 0 \\ -1, & i = j \neq 0 \\ 0, & \text{otherwise} \end{cases}$$

You can view η as a new metric defined on the Minkowski space, and whenever we mention the Minkowski space, the associated metric is assumed to be η.

There is another way to obtain equation 4. First, define space-time separation to be $(x^0)^2 - ((x^1)^2 + (x^2)^2 + (x^3)^2)$, where the terms in brackets represent the modulus of the vector in the space $\mathbb{R}^3$. Then shrink x close to 0, changing x^i to dx^i. Lorentz transformation preserves the proper time formulae, i.e. the space-time separation is preserved under Lorentz transformation when dx is small.

Preservation of space-time separation translates as $x' \cdot x' = x \cdot x$, using metric η, we have

$$x' \cdot x' = (x')^T \eta x' = x^T \Lambda^T \eta \Lambda x = x^T \eta x = x \cdot x$$

this holds for all x, so $\Lambda^T \eta \Lambda = \eta$. Conversely, if we take any linear map Λ satisfying $\Lambda^T \eta \Lambda = \eta$ and use it to transform the Minkowski space, space-time separation will be preserved. This defines the Lorentz group.

1.6 Lorentz Group

Definition 1 (Lorentz group)**.** The Lorentz group is defined by

$$O(1,3) = \{\Lambda \in GL(4, \mathbb{R}) : \Lambda^T \eta \Lambda = \eta\}$$

note here η is treated as a matrix. (even if it is essentially a bilinear form)

Of course, the Lorentz transformation in direction x discussed above is in the Lorentz group. Such transformation is called a boost. Any transformation to the frame of a moving object's(at constant speed) perspective is defined as a *boost*. Spatial rotations are also in the Lorentz group. (i.e. rotations that preserve x^0)

$O(1,3)$ is a group under operation: composition. Indeed, map composition is associative. All orthogonal maps have inverse and the multiplicative identity is given by the id map. The collection of all boosts, all rotations (not only the spatial ones) and their compositions form a subgroup of $O(1,3)$ called $SO^+(1,3)$.

Remark. O in $O(1,3)$ represents orthogonal group. Indeed, all transformation in the Lorentz group preserves vector norm, so they are orthogonal in the Minkowski space. SO (special orthogonal group) means that transformations in this group all have determinant 1. The determinant of a matrix represents the scaling of the area of a unit grid, (see *Essence of Linear algebra* chapter 6 on YouTube) So if transformation L has determinant 1, the distance travelled by any object between t_1, t_2 will not be changed. (this distance can be calculated by integration, essentially summing rectangular blocks) We call such transformations *Proper*. If transformation is improper, the distance travelled by an object becomes ambiguous.

The + sign in $SO^+(1,3)$ means that the direction of time is preserved, i.e. $\Lambda^0_0 > 0$. Such transformations are called *orthochronous*. If $\Lambda^0_0 < 0$, the transformation is *antichronous*. The collection of all transformations with $\Lambda^0_0 < 0$ is called $O^-(1,3)$ but it is not a subgroup of $O(1,3)$, because multiplicative identity id map is not in $O^-(1,3)$

Two simple examples lie in $O(1,3)$ but no in $SO^+(1,3)$: time-reversal $\mathcal{T} := \text{diag}(-1,1,1,1)$, although it is improper, the distance travelled is just reversed in sign. The parity transformation $\mathcal{P} := \text{diag}(1,-1,-1,-1)$ is also improper and it represents flipping the whole space like a mirror. Again, the distance travelled in just reversed in sign.

$\mathcal{T}, \mathcal{P}$ play important role because it can be shown that $O(1,3)$ has four pieces (of manifolds):

$$SO^+(1,3), \mathcal{T}SO^+(1,3), \mathcal{P}SO^+(1,3), \mathcal{P}\mathcal{T}SO^+(1,3)$$

the four manifolds are unconnected to each other, so $O(1,3)$ is an unconnected manifld.

It is not hard to see that the inverse metric tensor of η is inverse of itself (see the inverse metric tensor in section 2.7), i.e. $\eta\eta = I_4$. Suppose v is a four-vector (i.e. v^i are components of a vector x s.t. the transformation rule is the same as Lorentz transformations), by definition, $v^{\mu'} = \Lambda^{\mu'}_\nu v^\nu$. We can find dual v^* of this four-vector v using η, define $v^*_\mu = \eta_{\mu\nu} v^\nu$ (This is a lowering of the index, see section 2.7) And conversely, $\eta^{\mu\nu} v^*_\nu = v^\mu$, where $\eta^{\mu\nu} = \eta_{\mu\nu}$ because the inverse metric of η is η itself. In matrix form, these equations mean $\eta v = v^*$, $\eta v^* = v$.

Lorentz group has dimension 6 because there are 6 transformations required to be a basis. They are: rotations around x^1, x^2, x^3 axis respectively and boosts along x^1, x^2, x^3 axis respectively. If you want to know more properties of the Lorentz group, look up the Lorentz algebra.

General Lorentz transformations or Poincaré transformations enable the translation of the origin so they are more flexible. We have secretly used them once when studying the Twins paradox (I moved [frame B1] upwards so that the even B arrives at star S is at the origin)

Four vectors

An event is a point in the Minkowski space (space-time), and sometimes we want to describe the displacement (between two events) or the velocity of an object, which should be a tensor(or if you don't know tensors yet, just understand it as a vector for now) in the Minkowski space. There may be many choices of coordinates, and the tensor itself will not change, but the components may change. So we need to take out its components and specify the transformation rules. As you may have guessed, transformation rules (for change of coordinates) are Lorentz transformation and inverse Lorentz transformation:

$$x^{\mu'} = \Lambda^{\mu'}_\nu x^\nu, \quad x^\nu = (\Lambda^{-1})^\nu_{\mu'} x^{\mu'}$$

Such tensors are all called *four-vectors.*

Four-velocity:

$$u^\mu := \frac{dx^\mu}{d\tau}$$

where $\tau = t/\gamma$ is the proper time. (the name and use of symbol u is to distinguish four-velocity from the ordinary velocity) If we define the usual velocity vector in the space $\mathbb{R}^3$, i.e. $\boldsymbol{v} := (\dot{x} = dx/dt, \dot{y} = dy/dt, \dot{z} = dz/dt)^T$, then

$$u^0 = \frac{dx^0}{d\tau} = c\frac{dt}{d\tau} = c\gamma$$

$$\text{for}\mu \in \{1,2,3\},\ u^\mu = \frac{dx^\mu}{d\tau} = \frac{dx^\mu}{dt}\frac{dt}{d\tau} = \gamma\dot{x}^\mu$$

therefore, $u = \gamma(c, \boldsymbol{v})^T$.

Note, even if $\boldsymbol{v} = 0$, $u = \gamma c = c$ (because when $v = 0$, $\gamma = 1/\sqrt{1 - v^2/c^2} = 1$. Therefore, an object stationary in space still has non-zero energy, given by $E_0 := mc^2$ (yes! this is the famous *Mass–energy equivalence* $E = mc^2$ you know) where m is the mass of the object. Note for moving objects, this energy depends on speed v,

$$E = m\gamma c^2 = mc^2/\sqrt{1 - v^2/c^2}$$

But this formula does not work for photons (as $v = c$ makes denominator 0), so we will use four-momentum to define the energy of photons.

Four-momentum

$$p^\mu := mu^\mu$$

so using vector equation,

$$p = mu = m\begin{pmatrix}\gamma c\\ \gamma\boldsymbol{v}\end{pmatrix} = \begin{pmatrix}\gamma mc\\ \gamma m\boldsymbol{v}\end{pmatrix}$$

the second component $\gamma m\boldsymbol{v}$ is denoted as $\boldsymbol{p}$, and when $v \ll c$, $\boldsymbol{p} \approx m\boldsymbol{v}$, which is the momentum given in classical physics. The first term is quite interesting,

$$\begin{aligned}p^0 = \gamma mc = \frac{mc}{\sqrt{1 - \frac{v^2}{c^2}}} &= mc(1 - \frac{v^2}{c^2})^{-1/2}\\ &= mc(1 + \frac{1}{2}\frac{v^2}{c^2} + \cdots) \quad \text{binomial expansion}\\ &= mc + \frac{1}{2}m\frac{v^2}{c} + o((\frac{v}{c})^4)\end{aligned}$$

Therefore,

$$cp^0 = mc^2 + \frac{1}{2}mv^2 = E_0 + E_k \quad \text{where } E_k \text{ means kinetic energy}$$

we define the total energy of the object to be $E := E_0 + E_k$, then $p^0 = E/c$. The first entry of the four-momentum encodes the total energy of an object.

Note (Minkowski) norm of p is invariant under Lorentz transform,

$$p \cdot p = \gamma^2 m^2 c^2 - \gamma^2 m^2 |\boldsymbol{v}|^2 = \frac{1}{1 - (v/c)^2}m^2(c^2 - v^2) = m^2c^2$$

(because this norm is invariant under Lorentz transform for all four-vectors) Now using the new formulae we just derived above, $p = (E/c, \boldsymbol{p})$,

$$p \cdot p = m^2c^2 = \frac{E^2}{c^2} - |\boldsymbol{p}|^2, \quad \Rightarrow \quad E^2 = |\boldsymbol{p}|^2c^2 + m^2c^4$$

this is the general form of mass-energy equivalence. When $\boldsymbol{p} = 0$ (massive object is stationary), this takes the form $E = mc^2$ ($E = -mc^2$ is also possible, but negative energies are not considered in this course) When $m = 0$ (like photons), we have $E = |\boldsymbol{p}|c$.

As a final remark, the partial derivative functor is co-variant. And we can get its component using $\partial^\mu = \eta^{\mu\nu}\partial_\nu = (\partial_0, -\partial_1, -\partial_2, -\partial_3)$. This is essentially raising the index using the metric tensor as mentioned in section 2.7.

2 Introduction to Tensors

Tensors are widely used in General Relativity and quantum mechanics, quantum computing. You may see someone explaining that tensors are arrays with various dimensions, for example, matrices are 2-dimensional arrays. But arrays are only an arbitrary way of representing tensors, not the essence of tensors. This section will be mainly based on the video series *tensors for beginners* by Eigenchris on YouTube, but this will be a more compact, shorter introduction specially designed for students with a math background. If you struggle to understand any concept, please check the YouTube videos.

Tensors are defined so that they represent physical quantities in a coordinate-free way (i.e. after a tensor is defined, it is the same in any other choice of the coordinate system) If we define vectors (rank-1 tensors) by 1-dimensional arrays like (a, b, c), after changing the coordinate system, the array becomes different. Instead, we can define a set of base vectors $B = \{v_1, v_2, v_3\}$, represent vector $v = av_1 + bv_2 + cv_3$ using an array $[v]_B := (a, b, c)$. a, b, c are called components of a vector. And then, we specify rules for transforming the array representation when the coordinate system is changed. i.e. vectors can be defined as components under a chosen basis, and a set of transformation rules.

2.1 Vectors are tensors

Suppose we have two sets of base vectors (for now we define vectors as arrows, and base vectors are arrows in space that span a framework covering the whole space) the original basis $\{\boldsymbol{e}_1, \cdots, \boldsymbol{e}_n\}$ and the new basis $\{\tilde{\boldsymbol{e}}_1, \cdots, \tilde{\boldsymbol{e}}_n\}$. How are the two sets of base vectors related? From linear algebra, we know each new basis vector $\tilde{\boldsymbol{e}}_i$ can be represented by a linear combination of the original basis, i.e. $\sum_{j=1}^n F_{ji}\boldsymbol{e}_j$ where F_{kj} are scalars. This is called *forward transformation*. We can build matrix F, where (j, i) component is $F = (F_{ji})$. For example, if

$$\tilde{\boldsymbol{e}}_1 = 2\boldsymbol{e}_1 + 1\boldsymbol{e}_2, \quad \tilde{\boldsymbol{e}}_2 = -1\boldsymbol{e}_1 + 1\boldsymbol{e}_2$$

then

$$\begin{pmatrix}\tilde{\boldsymbol{e}}_1 & \tilde{\boldsymbol{e}}_2\end{pmatrix} = \begin{pmatrix}\boldsymbol{e}_1 & \boldsymbol{e}_2\end{pmatrix} \underbrace{\begin{pmatrix}2 & -1 \\ 1 & 1\end{pmatrix}}_{F}$$

All basis will be written as row vectors in this note.

For convenience, we use Einstein's notation for summations. In $\sum_{j=1}^n F_{ji}\boldsymbol{e}_j$, we omit the $\sum$ and representing the sum as $F_{ji}\boldsymbol{e}_j$. Note j here is just a placeholder, $F_{ki}\boldsymbol{e}_k$ will represent the same sum.

Now if we represent $\boldsymbol{e}_i$ by $B_{ki}\tilde{\boldsymbol{e}}_k$, this is the *backward transformation* and we can define matrix $B := (B_{ki})$.

As you may have guessed, F and B are inverses of each other

Proof. Starting from forward transformation,

$$\tilde{\boldsymbol{e}}_i = F_{ji}\boldsymbol{e}_j = F_{ji}(B_{kj}\tilde{\boldsymbol{e}}_k) = F_{ji}B_{kj}\tilde{\boldsymbol{e}}_k$$

so $B_{kj}F_{ji} = \delta_{ik}$ where the Kronecker delta δ_{ik} is defined by

$$\delta_{ik} = \begin{cases} 1, & i = k \\ 0, & i \neq k \end{cases}$$

$B_{kj}F_{ji}$ is actually the (k, i) entry of matrix BF, so $B_{kj}F_{ji} = \delta_{ik}$ simply means BF is identity matrix.

You can start with backward transformation $\boldsymbol{e}_i = B_{ki}\tilde{\boldsymbol{e}}_k$, and plug the forward transformation to prove $FB = I$. □

Now given any other vector v, we can represent it using the new and original basis,

$$v = v_i\boldsymbol{e}_i, \quad v = \tilde{v}_i\tilde{\boldsymbol{e}}_i$$

$v_i, \tilde{v}_i$ are called *vector components* and if they are arranged into a 1-dimensional array, we have the array representation of a vector. Now we study how to use forward and backward transformation to find relationships between v_i and $\tilde{v}_i$.

From new basis to old basis:

$$v = \tilde{v}_i \tilde{\boldsymbol{e}}_i = \tilde{v}_i F_{ji} \boldsymbol{e}_j$$

comparing this with $v = v_i \boldsymbol{e}_i$, we have $v_j = F_{ji}\tilde{v}_i$. Similarly, one can deduce $\tilde{v}_j = B_{ji} v_i$. So forward transformation brings us from vector components under the new basis back to the old basis, even though forward transformation takes an old base vector to a new base vector. Therefore, we say vector components are *contravariant.* And for this reason, vector components are written as v^i instead of v_i. (Exponents will not appear in this note, so whenever you see v^i, it means a contravariant component)

Now you can see why vectors are not just 1-d arrays. Because vector components change with the choice of basis.

2.2 Co-vectors are tensors

Co-vectors are represented by row vectors, but they are in the essence different to vectors. Because co-vectors are *covariant* tensors whereas vectors are *contra-variant* tensors. We will explain this later.

Co-vectors are functions acting on vectors, for example, $\begin{pmatrix} 1 & 2 \end{pmatrix}$ acts on vectors as follows

$$\begin{pmatrix} 1 & 2 \end{pmatrix} \begin{pmatrix} 5 \\ 4 \end{pmatrix} = 1 \times 5 + 2 \times 4 = 13$$

and in general,

$$\begin{pmatrix} 1 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = x + 2y$$

The strict mathematical definitions of co-vectors are linear functions $\alpha : V \to \mathbb{R}$ (where V is a vector space) They form a dual space V^* of V. We can find a basis for this dual space based on the given basis $\{\boldsymbol{e}_i\}$ of V:

$$\boldsymbol{\epsilon}^i(\boldsymbol{e}_j) := \delta_{ij}$$

this is called the *dual basis* of $\{\boldsymbol{e}_i\}$. (you will see why indices of dual basis are written above) So each $\boldsymbol{\epsilon}^i$ will only keep values along $\boldsymbol{e}_i$, i.e. if given vector $v = v^i \boldsymbol{e}_i$,

$$\begin{aligned} \boldsymbol{\epsilon}^i(v) &= \boldsymbol{\epsilon}^i(v^j \boldsymbol{e}_j) \quad \text{be careful when using Einstein's notation, sometimes you need to give indices new names to avoid conflict} \\ &= v^j \boldsymbol{\epsilon}^i(\boldsymbol{e}_j) \quad \text{by linearity} \\ &= v^j \delta_{ij} \quad \text{by definition of } \boldsymbol{\epsilon}^i \\ &= v^i \end{aligned}$$

Proposition 2.1. *The defined $\{\boldsymbol{\epsilon}^i\}$ form a basis of V^**

Proof. Given any $\alpha \in V^*$,

$$\alpha(v) = \alpha(v^i \boldsymbol{e}_i) = v^i \alpha(\boldsymbol{e}_i) = \alpha(\boldsymbol{e}_i)\boldsymbol{\epsilon}^i(v)$$

therefore, we can define components of α by $\alpha_i := \alpha(\boldsymbol{e}_i)$. Then, $\alpha = \alpha_i \boldsymbol{\epsilon}^i$. The uniqueness of these components is left as an exercise. □

Co-vectors seem to be the same as vectors until now, but transformation rules make them different from vectors. We can construct another dual basis $\{\tilde{\boldsymbol{\epsilon}}_i\}$ using new basis $\{\tilde{\boldsymbol{e}}_i\}$ and decompose any $\alpha \in V^*$ into $\alpha = \tilde{\alpha}_i \tilde{\boldsymbol{\epsilon}}_i$ by defining $\tilde{\alpha}_i := \alpha(\tilde{\boldsymbol{e}}_i)$. $\tilde{\alpha}_i$ are components of α under the new dual basis. Now we find a way to transform between old and new components:

$$\begin{aligned} \tilde{\alpha}_i &= \alpha(\tilde{\boldsymbol{e}}_i) \\ &= \alpha(F_{ji} \boldsymbol{e}_i) \quad \text{using forward transformation} \\ &= F_{ji} \alpha(\boldsymbol{e}_i) = F_{ji} \alpha_i \end{aligned}$$

so forward transformation brings us from components on the old dual basis to the new dual basis. Similarly one will find $\alpha_i = B_{ji}\tilde{\alpha}_i$. Therefore, co-vectors are called *covariant*. Transformation rules for co-vector components are in opposite direction to rules for vectors. Whenever we have *covariant* components, we write the index at the bottom, e.g. α_i.

2.3 Linear maps are tensors

Linear maps are also tensors. We represent linear maps as 2d-arrays (or matrices) but again, the choice of basis changes the matrix representation, so we need to understand the intrinsic meaning of linear maps. Geometrically, if we draw grids on the space, linear maps are transformations that keep the grid lines parallel, evenly-spaced and will not move the origin. (If you are confused about imaging these grid lines, check out 3Blue1Brown's videos series *Essence of Linear Algebra*) From linear algebra, we know linear maps are defined as functions $L : V \to W$ (where V, W are two vector spaces) s.t. L is linear. i.e. $L(v + w) = L(v) + L(w)$ and $L(kv) = kL(v)$ where $k \in \mathbb{R}$.

For this section, we assume that $L : V \to V$ (i.e. domain and codomain are the same). Given vector $v = v^i \boldsymbol{e}_i$,

$$L(v) = L(v^i \boldsymbol{e}_i) = v^i L(\boldsymbol{e}_i)$$

note here $L(\boldsymbol{e}_i)$ are vectors, so for each i, we can write $L(\boldsymbol{e}_i) = L_i^j \boldsymbol{e}_j$. The reason why j is at the top will become clear after explaining transformation rules for linear maps.

Using L_i^j, we can represent linear maps in a matrix, by putting L_i^j at ith column and jth row. For example, when $V = \mathbb{R}^2$, a matrix representation of L looks like

$$\begin{pmatrix} L_1^1 & L_2^1 \\ L_1^2 & L_2^2 \end{pmatrix}$$

Note ith entry of $L(v)$ is ith entry of the sum $v^j L(\boldsymbol{e}_j)$, which is $v^j L_j^i$. You may recognise this is how the matrix (L_j^i) is multiplied with vectors. So matrix multiplication rule can be deduced from these abstract ideas.

Suppose we have $[v]_{\boldsymbol{e}_i}$, the array representation of v in the old basis $\{\boldsymbol{e}_i\}$, and $[v]_{\tilde{\boldsymbol{e}}_i}$, the array representation of v in the new basis. And suppose $L_{\boldsymbol{e}_i} := (L_j^i)$ is the matrix representation of L under old basis, then by the above arguments, $[Lv]_{\boldsymbol{e}_i} = L_{\boldsymbol{e}_i}[v]_{\boldsymbol{e}_i}$. But clearly the matrix $L_{\boldsymbol{e}_i}$ cannot be used on the new basis. Suppose $L_{\tilde{\boldsymbol{e}}_i} := (\widetilde{L_i^j})$ is the matrix representation of L under the new basis, we study how to find $\widetilde{L_i^j}$ from L_i^j. i.e. study transformation rules for linear maps.

The idea (see figure 13) is that instead of going directly from $[v]_{\tilde{\boldsymbol{e}}_i}$ to $[Lv]_{\tilde{\boldsymbol{e}}_i}$, we first use forward transformation F to find components in the old basis $[v]_{\boldsymbol{e}_i}$ (remember vector components are contravariant), then apply L to get $[Lv]_{\boldsymbol{e}_i}$. Finally, we use backward transformation to find components of Lv in the new basis, i.e. $[Lv]_{\tilde{\boldsymbol{e}}_i}$. So from this process, you already see that $L_{\tilde{\boldsymbol{e}}_i} = BL_{\boldsymbol{e}_i}F$.

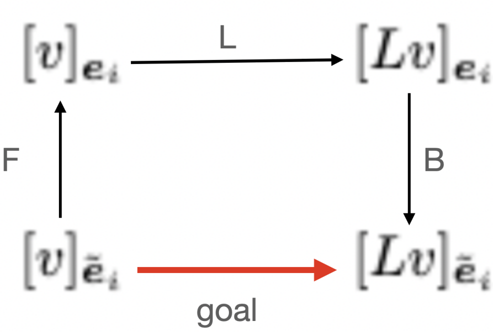


Figure 13: Transformation rule for linear maps

Another way to get the transformation rule is using Einstein's notation, this is left as an exercise. But before trying that, you may find it convenient to write F_i^j instead of F_{ij} for forward transformation, and B_k^l instead of

B_{kl}. Because if matrices are written in this way, the (k, i) entry of AB would be $A^k_j B^j_i$. i.e. lower index of the first one matches the top index of the second one.

For the linear map transformation rule, you should get

$$\widetilde{L^l_i} = B^l_k L^k_l F^j_i \tag{5}$$

which basically means $L_{\tilde{e}_i} = BL_{e_i}F$. Now based on equation 5, we can derive a way to go from $\widetilde{L^l_i}$ to L^k_l.

$$\begin{aligned}
\widetilde{L^l_i} &= B^l_k L^k_l F^j_i \\
\Rightarrow \quad F^s_l \widetilde{L^l_i} B^i_t &= F^s_l B^l_k L^k_l F^j_i B^i_t \quad \text{multiply F, B on both sides} \\
\Rightarrow F^s_l \widetilde{L^l_i} B^i_t &= \delta_{sk} L^k_l \delta_{jt} \quad \text{because } F, B \text{ are inverses of each other} \\
\Rightarrow F^s_l \widetilde{L^l_i} B^i_t &= L^s_t \quad \text{because RHS is only non-zero when } j = t,\ s = k
\end{aligned}$$

That gives the backward transformation rule for linear maps.

Summary
Vector components and dual basis transform under the contra-variant law:

$$\tilde{\boldsymbol{\epsilon}}^i = B^i_j \boldsymbol{\epsilon}^j, \quad \boldsymbol{\epsilon}^i = F^i_j \tilde{\boldsymbol{\epsilon}}^j$$

$$\tilde{v}^i = B^i_j v^j, \quad v^i = F^i_j \tilde{v}^j$$

so they are called (1, 0) tensors. (1, 0) simply means there is one contra-variant law used and no co-variant law.

Co-vector components and vector basis transform under the co-variant law:

$$\tilde{\boldsymbol{e}}_j = F^i_j \boldsymbol{e}_i, \quad \boldsymbol{e}_j = B^i_j \tilde{\boldsymbol{e}}_i$$

$$\tilde{\alpha}_j = F^i_j \alpha_i, \quad \alpha_j = B^i_j \tilde{\alpha}_i$$

as you may have guessed, they are called (0, 1) tensors, with no contra-variant law and one co-variant law.

Finally, linear maps are (1, 1) tensors following transformation rules:

$$\widetilde{L^i_j} = B^i_k L^k_l F^j_l, \quad L^i_j = F^i_k \widetilde{L^k_l} B^l_j$$

transformations use one co-variant and one contra-variant.

2.4 Metric Tensor

Metric tensors are defined to measure the length of vectors, and then, inner products.

We already know that Pythagoras' theorem will not work on a non-orthogonal basis. So the inner product is used to define the length of a vector instead, i.e. $\sqrt{v \cdot v}$. Suppose $v = v^i \boldsymbol{e}_i$, we can expand the expression for the inner product by

$$v \cdot v = v^i \boldsymbol{e}_i v^j \boldsymbol{e}_j = v^i v^j (\boldsymbol{e}_i \cdot \boldsymbol{e}_j)$$

In another word, if we manage to figure out $\boldsymbol{e}_i \cdot \boldsymbol{e}_j$ for all i, j, we can find the length of the vector under this basis. So define metric tensor components $g_{ij} := \boldsymbol{e}_i \cdot \boldsymbol{e}_j$. It can be written into a matrix $G := (g_{ij})$, and then

$$v \cdot v = [v]^T_{\boldsymbol{e}_i} G [v]_{\boldsymbol{e}_i}$$

If v is dotted with a different vector, $v \cdot w = v^i w^j g_{ij}$. So metric tensor enables us to find any inner product under this basis. Note by the symmetry of the inner product, the matrix G will always be symmetric.

Now we study the transformation rule, suppose $\widetilde{g_{ij}} := \tilde{\boldsymbol{e}}_i \cdot \tilde{\boldsymbol{e}}_j$ is the metric tensor under the new basis, then

$$\begin{aligned}
\widetilde{g_{ij}} &= \tilde{\boldsymbol{e}}_i \cdot \tilde{\boldsymbol{e}}_j \quad \text{definition} \\
&= F_i^k e_k F_j^l e_l = F_i^k F_j^l (e_k \cdot e_l) \\
&= F_i^k F_j^l g_{kl} \quad \text{by definition of } g_{kl}
\end{aligned}$$

so two forward transformations are required. One can follow similar process and obtain $g_{kl} = B_k^i B_l^j \widetilde{g_{ij}}$. Since two co-variant transformation rules are used, metric tensors are called (0, 2) tensors.

Proposition 2.2. *The length of the vector defined above is well-defined. i.e. it is the same under all basis*

Proof. We start from the formulae of the inner product on the new basis

$$\begin{aligned}
v \cdot v &= \tilde{v^i} \tilde{v^j} \widetilde{g_{ij}} \quad \text{formulae in the new basis} \\
&= (B_a^i v^a)(B_b^j v^b)(F_i^k F_j^l g_{kl}) \quad \text{by transformation rules} \\
&= v^a v^b g_{kl} F_i^k B_a^i F_j^l B_b^j \\
&= v^a v^b g_{kl} \delta_{ka} \delta_{lb} \quad \text{as } F, B \text{ are inverses to each other} \\
&= v^k v^l g_{kl}
\end{aligned}$$

The final line is the formulae of vector length in the old basis, so we have proved the proposition. □

If we set g_{ij} free, instead of requiring them to be inner products of $\boldsymbol{e}_i$, we get bilinear forms. Bilinear forms are functions $g : V \times V \to \mathbb{R}$ defined by $g(v, w) = v^i w^j g_{ij}$. You may check that bilinear forms are also (0, 2) tensors. This time if we build matrix $G := (g_{ij})$, it may not be symmetric. But the matrix representation of a metric tensor being symmetric does not necessarily mean all symmetric matrices represent a metric tensor. For example, consider

$$\begin{pmatrix} 1 & -5 \\ -5 & 1 \end{pmatrix}$$

if we plug in $\begin{pmatrix} 1 & 1 \end{pmatrix}$,

$$\begin{pmatrix} 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & -5 \\ -5 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = -8 < 0$$

this will never happen for metric tensors because the "length" of a vector cannot be negative.

2.5 A first definition

After seeing many examples, hopefully, you have a feeling of what a tensor should be. Now we define a general (m, n) tensor, it is something that is invariant under a change of coordinates, and its components $T^{ijk\cdots}_{rst\cdots}$ transform under the rule

$$\widetilde{T^{ijk\cdots}_{rst\cdots}} = \underbrace{B_i^a B_j^b B_k^c \cdots}_{m \text{ contra-variant}} \; T^{ijk\cdots}_{rst\cdots} \; \underbrace{F_x^r F_y^s F_t^k \cdots}_{n \text{ co-variant}}$$

2.6 Constructing tensors

The definition of tensor above is still vague, what is "something"? A geometrical object? An arrow? Now we introduce a new way of viewing tensors. Essentially, vectors and co-vectors and be combined, just like we can build a matrix by multiplying a row vector with a column vector (the outer product). The problem of the outer product is that some matrices are not the outer product of vectors, e.g. you will never find a, b, c, d s.t.

$$\begin{pmatrix} a \\ b \end{pmatrix} \begin{pmatrix} c & d \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 2 & 3 \end{pmatrix}$$

and the reason is that outer products only have rank 1, it will not be able to construct matrices of higher rank.

But we can try to combine two bases $\{\boldsymbol{e}_i\}, \{\boldsymbol{\epsilon}_i\}$. Note the matrices E_{ij} defined to be a matrix with 1 at entry (i, j) and 0 otherwise. Such matrices form a basis for set of all linear maps. For example, the bilinear form $G = g_{ij} E_{ij}$.

Further, each E_{ij} is outer product $\boldsymbol{e}_j\boldsymbol{\epsilon}^i$. Therefore, $\{\boldsymbol{e}_j\boldsymbol{\epsilon}^i\}$ is a basis. So a linear map can be written as $L = L^i_j\boldsymbol{e}_i\boldsymbol{\epsilon}^j$.

When acting on vectors, $Lv = (L^i_j\boldsymbol{e}_i\boldsymbol{\epsilon}^j)(v^k\boldsymbol{e}_k)$. Here co-vectors can act on vectors, yielding $L^i_jv^k\boldsymbol{e}_i\boldsymbol{\epsilon}^j(\boldsymbol{e}_k) = L^i_jv^k\boldsymbol{e}_i\delta_{jk} = L^i_jv^j\boldsymbol{e}_i$. (And this indeed aligns with the matrix multiplication rule) Again, other bases may be chosen, so the linear map L may also have representation $L = \widetilde{L^i_j}\tilde{\boldsymbol{e}}_i\tilde{\boldsymbol{\epsilon}}^j$. Now the transformation rules can be derived directly from those of vectors and co-vectors, e.g.

$$L = L^k_l\boldsymbol{e}_k\boldsymbol{\epsilon}^l = L^k_l(B^i_k\tilde{\boldsymbol{e}}_i)(F^l_j\tilde{\boldsymbol{\epsilon}}^j) = B^i_kL^k_lF^l_j(\tilde{\boldsymbol{e}}_i\tilde{\boldsymbol{\epsilon}}^j)$$

so $\widetilde{L^i_j} = B^i_kL^k_lF^l_j$, same as the formulae derived before.

Multiplication between $\boldsymbol{e}_j$ and $\boldsymbol{\epsilon}^i$ above is actually a tensor product. Before rigorously defining tensor product, let us look at another example: combining co-vector with co-vector. Defining bilinear form $B := B_{ij}\boldsymbol{\epsilon}^i\boldsymbol{\epsilon}^j$, we see that it is capable of taking two vectors (because there are two co-vectors), just like bilinear forms. Let us try find $B(v, w)$

$$\begin{aligned} B(v,w) &= B_{ij}\boldsymbol{\epsilon}^i\boldsymbol{\epsilon}^j(v^k\boldsymbol{e}_k, w^l\boldsymbol{e}_l) \\ &= B_{ij}\boldsymbol{\epsilon}^i(v^k\boldsymbol{e}_k)\boldsymbol{\epsilon}^j(w^l\boldsymbol{e}_l) \\ &= B_{ij}v^kw^l\boldsymbol{\epsilon}^i(\boldsymbol{e}_k)\boldsymbol{\epsilon}^j(\boldsymbol{e}_l) \\ &= B_{ij}v^kw^l\delta_{ik}\delta_{jl} \\ &= B_{ij}v^iw^j \end{aligned}$$

One can follow similar procedures and recover the transformation rule of bilinear forms: $\widetilde{B_{ij}} = F^k_iF^l_jB_{kl}$.

Now we introduce the tensor product of two arrays, the Kronecker product of $A \otimes B$ is essentially distributing A to each entry of B, for example, the following is a linear map defined by a vector and a co-vector:

$$\begin{pmatrix} v^1 \\ v^2 \end{pmatrix} \otimes \begin{pmatrix} \alpha_1 & \alpha_2 \end{pmatrix} = \begin{pmatrix} \alpha_1\begin{pmatrix} v^1 \\ v^2 \end{pmatrix} & \alpha_2\begin{pmatrix} v^1 \\ v^2 \end{pmatrix} \end{pmatrix} = \begin{pmatrix} \begin{pmatrix} \alpha_1v^1 \\ \alpha_1v^2 \end{pmatrix} & \begin{pmatrix} \alpha_2v^1 \\ \alpha_2v^2 \end{pmatrix} \end{pmatrix}$$

it is a row vector of column vectors. When multiplied with a vector,

$$\begin{pmatrix} \begin{pmatrix} \alpha_1v^1 \\ \alpha_1v^2 \end{pmatrix} & \begin{pmatrix} \alpha_2v^1 \\ \alpha_2v^2 \end{pmatrix} \end{pmatrix}\begin{pmatrix} w^1 \\ w^2 \end{pmatrix} = w^1\begin{pmatrix} \alpha_1v^1 \\ \alpha_1v^2 \end{pmatrix} + w^2\begin{pmatrix} \alpha_2v^1 \\ \alpha_2v^2 \end{pmatrix}$$

a linear combination of columns. This may seem redundant, but when we use it on bilinear forms (a covector-covector pair):

$$\begin{aligned} &\begin{pmatrix} \alpha_1 & \alpha_2 \end{pmatrix} \otimes \begin{pmatrix} \beta_1 & \beta_2 \end{pmatrix} \\ &\qquad = \begin{pmatrix} \beta_1\begin{pmatrix} \alpha_1 & \alpha_2 \end{pmatrix} & \beta_2\begin{pmatrix} \alpha_1 & \alpha_2 \end{pmatrix} \end{pmatrix} \\ &\qquad = \begin{pmatrix} \begin{pmatrix} \alpha_1\beta_1 & \alpha_2\beta_1 \end{pmatrix} & \begin{pmatrix} \alpha_1\beta_2 & \alpha_2\beta_2 \end{pmatrix} \end{pmatrix} \end{aligned}$$

Compare the following ways to find $B(v, w)$ (where components of B are B_{ij} for $i, j = 1, 2$)

$$B(v,w) = \begin{pmatrix} v^1 & v^2 \end{pmatrix}\begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}\begin{pmatrix} w^1 \\ w^2 \end{pmatrix}$$

v, w should all be vectors, but we suddenly flip v into row vector even though v is not a co-vector. Now using the Kronecker product:

$$\begin{aligned} B(v,w) &= \begin{pmatrix} \begin{pmatrix} B_{11} & B_{12} \end{pmatrix} & \begin{pmatrix} B_{21} & B_{22} \end{pmatrix} \end{pmatrix}\begin{pmatrix} v^1 \\ v^2 \end{pmatrix}\begin{pmatrix} w^1 \\ w^2 \end{pmatrix} \\ &= \begin{pmatrix} \begin{pmatrix} B_{11} & B_{12} \end{pmatrix}v^1 + \begin{pmatrix} B_{21} & B_{22} \end{pmatrix}v^2 \end{pmatrix}\begin{pmatrix} w^1 \\ w^2 \end{pmatrix} \\ &= \begin{pmatrix} \begin{pmatrix} B_{11}v^1 + B_{21}v^2 & B_{12}v^1 + B_{22}v^2 \end{pmatrix} \end{pmatrix}\begin{pmatrix} w^1 \\ w^2 \end{pmatrix} \\ &= (B_{11}v^1 + B_{21}v^2)w^1 + (B_{12}v^1 + B_{22}v^2)w^2 \\ &= B_{11}v^1w^1 + B_{21}v^2w^1 + B_{12}v^1w^2 + B_{22}v^2w^2 \end{aligned}$$

this feels more natural.

Kronecker product combines two arrays and forms a new array, and the tensor product combines two tensors forming a new tensor. You will find these products highly related. Now suppose we have a linear map defined by tensor product, $\boldsymbol{e}_i \otimes \boldsymbol{\epsilon}^j$, it acts on vectors as follows

$$\begin{aligned}(\boldsymbol{e}_i \otimes \boldsymbol{\epsilon}^j)(v) &= \boldsymbol{e}_i \otimes \boldsymbol{\epsilon}^j(v) \\ &= \boldsymbol{e}_i \otimes \boldsymbol{\epsilon}^j(v^k \boldsymbol{e})k) \\ &= v^k \boldsymbol{e}_i \boldsymbol{\epsilon}^j(\boldsymbol{e}_k) \\ &= v^k \boldsymbol{e}_i \delta jk \\ &= v^j \boldsymbol{e}_i\end{aligned}$$

Comparison between Kronecker product and tensor product

- Tensor product
$$v \otimes \alpha = (v^i \boldsymbol{e}_i) \otimes (\alpha_j \boldsymbol{\epsilon}^j) = v^i \alpha_j (\boldsymbol{e}_i \otimes \boldsymbol{\epsilon}^j)$$
- Kronecker product
$$\begin{pmatrix} v^1 \\ v^2 \end{pmatrix} \otimes \begin{pmatrix} \alpha_1 & \alpha_2 \end{pmatrix} = \begin{pmatrix} \begin{pmatrix} \alpha_1 v^1 \\ \alpha_1 v^2 \end{pmatrix} & \begin{pmatrix} \alpha_2 v^1 \\ \alpha_2 v^2 \end{pmatrix} \end{pmatrix}$$

if you treat $\boldsymbol{e}_i \otimes \boldsymbol{\epsilon}^j$ as basis matrices E_{ij}, then these two products are essentially doing the same thing.

Finally, we can define tensor using the tensor product.

Definition 2 (Tensors)**.** Tensors are collections of vectors and co-vectors combined using the tensor product.

For convenience, $\otimes$ will be omitted. So $\boldsymbol{e}_i \otimes \boldsymbol{\epsilon}^j$ will be written as $\boldsymbol{e}_i \boldsymbol{\epsilon}^j$. For example, linear maps are constructed using a pair of vectors and co-vector, and bilinear forms are constructed using a pair of co-vectors. But we can construct higher rank tensors easily using this definition, for example, the (2, 1) tensor

$$Q := Q^{ij}_k \boldsymbol{e}_i \boldsymbol{e}_j \boldsymbol{\epsilon}^k$$

the transformation rule can be easily figured out by using transformation rules on the base vectors/co-vectors. If you want to figure out the shape of the array representation of Q, use the Kronecker product $v \otimes w \otimes \alpha$ where v, w are vectors and α is co-vector.

One small remark is that tensors acting on tensors may not be well defined, e.g.

$$\boldsymbol{\epsilon}^i \boldsymbol{\epsilon}^j (\boldsymbol{e}_a, \boldsymbol{e}_b)$$

there are four ways to define this:

$$\boldsymbol{\epsilon}^i(\boldsymbol{e}_a)\boldsymbol{\epsilon}^j(\boldsymbol{e}_b), \quad \boldsymbol{\epsilon}^i(\boldsymbol{e}_b)\boldsymbol{\epsilon}^j(\boldsymbol{e}_a), \quad \boldsymbol{\epsilon}^i(\boldsymbol{e}_a)\boldsymbol{e}_b, \quad \boldsymbol{\epsilon}^j(\boldsymbol{e}_b)\boldsymbol{e}_a$$

there could be more ways to define this. Similarly, there is no unique way to multiply arrays if they are complicated.

There is a strict mathematical definition of tensors product, it is a map taking two tensors v, α, yielding another tensor, and satisfies the scaling and adding rules shown below

- $n(v\alpha) = (nv)\alpha = v(n\alpha)$ for any $n \in \mathbb{R}$
- $v\alpha + v\beta = v(\alpha + \beta)$
- $v\alpha + w\alpha = (v + w)\alpha$

i.e. the tensor product is a bilinear map on the tensors.

Denote the dual vector space by V^* (a space consisting of all the co-vectors on V) We can form new spaces like $V \otimes V^*$ (space of vector-covector pairs, i.e. (1, 1) tensors), $V^* \otimes V^*$ (space of covector-covector pairs, i.e. (0, 2) tensors). But you can essentially take tensor products of an arbitrary number of vector spaces. Such spaces are all called *tensor product*, and vector spaces/dual vector spaces are special cases of tensor spaces.

Each tensor space is a module in the sense that it can act on some rings like V, V^*, $V \times V$. (this is NOT a tensor product, simply ordered pairs of vectors) For example, given bilinear form $B_{ij}\boldsymbol{\epsilon}^i\boldsymbol{\epsilon}^j$, it can act on

- V by $B(v) = B_{ij}v^i =: \alpha_j$, i.e. $V \to V^*$
- V by $B(w) = B_{ij}w^j =: \beta_i$, i.e. $V \to V^*$ (different from the previous one, because the first index of B is preserved instead of the second one)
- $V \times V$ by $B(v, w) = B_{ij}v^i w^j =: s \in \mathbb{R}$, i.e. $V \times V \to \mathbb{R}$. This is the bilinear form that we know.

2.7 Lowering and raising of indices

You have seen above that bilinear forms actually allow us to travel from $V \to V^*$. And metric tensor, the special case of bilinear form, allows us to "lower" the index. Given a vector $v = v^i\boldsymbol{e}_i$, we wish to find a corresponding co-vector $v_i\boldsymbol{\epsilon}^i$. A first idea may be defining correspondence $\boldsymbol{e}_i \mapsto \boldsymbol{\epsilon}^i$ and then $v_i = v^i$. But this is not quite right because base vectors transform using co-variant rules and co-vectors transform under contra-variant rules. For example, if $\tilde{\boldsymbol{e}}_i = 2\boldsymbol{e}_i$, then $\tilde{\boldsymbol{\epsilon}}^i = 1/2\boldsymbol{\epsilon}^i$. Then the correspondence in the new coordinate system must be $\tilde{\boldsymbol{e}}_i \mapsto 4\tilde{\boldsymbol{\epsilon}}^i$. So the correspondence change with coordinates, which is not preferable. Therefore, metric tensors are used. Essentially, we define correspondence $v \mapsto \alpha$ where $\alpha \in V^*$ is defined by

$$\alpha(w) = v \cdot w = g_{ij}v^i w^j = g_{ij}v^i\boldsymbol{\epsilon}_j(w)$$

so as you can see from the above equation, components of α are $\alpha_i := g_{ji}v^j$. We have found that $v_j = \alpha_j = g_{ji}v^j$. If a new coordinate system is used, this rule remains unchanged: $\tilde{v}_j = \widetilde{g_{ji}}\tilde{v}^j$.

Warning Please do not assume that $v_i = v^i$, they are only identical when $g_{ij} = \delta_{ij}$ (i.e. in orthonormal basis)

Now, we study a type of tensor called an inverse metric tensor, living in side $V \otimes V$ (compared to $V^* \otimes V^*$ for metric tensors), that can raise index. i.e. Given co-vector $\alpha_i\boldsymbol{\epsilon}^i$, we can find a corresponding vector $\alpha^i\boldsymbol{e}_i$ in V. We define components $\mathfrak{g}^{ki}$ of inverse of $g_{ij}\boldsymbol{\epsilon}^i\boldsymbol{\epsilon}^j$ to be s.t. $\mathfrak{g}^{ki}g_{ij} = \delta^k_j$.

Take the index lowering equation $v_i = g_{ij}v^j$ and multiply inverse of g_{ij} on both sides:

$$\mathfrak{g}^{ki}v_i = \mathfrak{g}^{ki}g_{ij}v^j = \delta^k_j v^j = v^k$$

i.e. $\mathfrak{g}^{ki}$ raises the indices of v_i. It is left as an exercise to check that this raising equation remains the same under a change of coordinates. i.e. if we are using $\tilde{\boldsymbol{\epsilon}}^i$ as dual basis, $\tilde{v}^k = \tilde{\mathfrak{g}}^{ki}\tilde{v}^i$ where $\tilde{\mathfrak{g}}^{ki}$ is defined so that $\tilde{\mathfrak{g}}^{ki}g_{ij} = \delta^k_i$.

Using the metric tensor and inverse metric tensor, you can lower/raise the index of any tensor. For example, $B_{ij}\mathfrak{g}^{jk} = B^k_i$. This changes the bilinear form to a linear map! You can also lower/raise multiple indices.

Lowering indices is sometimes represented by $\flat$ symbol from music. For example, if $v = v^i\boldsymbol{e}_i$, then $\flat v = v_j\boldsymbol{\epsilon}^j$. And raising index is represented by $\sharp$. If $v = v_i\boldsymbol{\epsilon}^i$, then $\sharp v = v^j\boldsymbol{e}_j$.

3 Tensor Calculus

One key difference between special relativity and general relativity is that the metric tensor η is constantly $(1, -1, -1, -1)$ but this is not the case in general relativity. The metric tensor depends on space-time (x, y, z, t). Therefore, studying what a tensor field is and how tensors change in the fields is essential. Therefore, this chapter will be dedicated to tensor calculus.

Tensor calculus is somehow an extension of multivariate calculus, so recall the following concepts and review if you cannot remember

- For set $S \subseteq \mathbb{R}^n$, a scalar field is $f : S \to \mathbb{R}$, a vector field is $\boldsymbol{f} : S \to \mathbb{R}^n$

- Partial derivative $\frac{\partial f}{\partial x_i}$ for $i = 1, \cdots, n$
- Chain rule: if $f(\boldsymbol{x}) = f(x_1, \cdots, x_n)$ and $x_i = g_i(t)$ for some function g_i, then

$$\frac{df}{dt} = \frac{\partial f}{\partial x_i}\frac{dx_i}{dt}$$

  (Eisenstein's notation is used)
- Gradient $\nabla_{\boldsymbol{x}} f(\boldsymbol{x})$, measures direction of steepest descent
- Directional derivative $\nabla_{\boldsymbol{v}} f = D_{\boldsymbol{v}} f := \boldsymbol{v} \cdot \nabla f$.
- Differential

$$df = \frac{\partial f}{\partial x_i} dx_i$$

- Cartesian and polar coordinate systems, and conversion between them

$$x = r\cos\theta,\ y = r\sin\theta \quad r = \sqrt{x^2 + y^2},\, \theta = \arctan\frac{y}{x}$$

Suppose a curve is traced by $R(t) = (x_1(t), \cdots, x_n(t))$, the arc length can be found by cutting the curve into infinitely many small line segments. The formula is

$$\int \left\| \frac{dR}{dt} \right\| dt$$

where dR/dt represents the line segments. Note $\|\boldsymbol{x}\| = \sqrt{\boldsymbol{x} \cdot \boldsymbol{x}}$, so arc length is

$$\int \sqrt{\left(\frac{\partial R}{\partial x_i}\frac{dx_i}{dt}\right) \cdot \left(\frac{\partial R}{\partial x_j}\frac{dx_j}{dt}\right)}\, dt = \int \sqrt{\frac{dx_i}{dt}\frac{dx_j}{dt}\left(\frac{\partial R}{\partial x_i} \cdot \frac{\partial R}{\partial x_j}\right)}\, dt$$

Usually, we pick the basis in a coordinate system along the coordinate lines (e.g. square grids for Cartesian coordinates) so for Cartesian,

$$\frac{\partial R}{\partial x_i} = \boldsymbol{e}_{x_i}$$

and for polar coordinates,

$$\frac{\partial R}{\partial r} = \boldsymbol{e}_r,\ \frac{\partial R}{\partial \theta} = \boldsymbol{e}_\theta$$

base vectors for polar coordinates change with the point picked. These bases are not orthonormal because $\|\partial R/\partial\theta\| \neq 1$. You may fix this by adding a scaling constant in front.

Using the above idea, we can rewrite forward and backward transformations using partial derivatives. Suppose we have two coordinate systems $(c_1, \cdots, c_n)$ and $(p_1, \cdots, p_n)$, and their basis vectors are

$$\boldsymbol{e}_i = \frac{\partial R}{\partial c_i},\ \tilde{\boldsymbol{e}}_i = \frac{\partial R}{\partial p_i}$$

By the chain rule,

$$\tilde{\boldsymbol{e}}_i = \frac{\partial R}{\partial p^i} = \frac{\partial R}{\partial c^j}\frac{\partial c^j}{\partial p^i} = \frac{\partial c^j}{\partial p^i}\boldsymbol{e}_j$$

so the forward transformation components $F_i^j = \partial c_j/\partial p_i$. Similarly, one can find the backward transformation is given by

$$\boldsymbol{e}_i = \frac{\partial p^j}{\partial c^i}\tilde{\boldsymbol{e}}_j$$

note here forward and backward transformations are both changing with coordinates. The transformations are different for different points in the space.

3.1 Derivatives are tensors

The derivative dR/dt (or the tangent vector) of $R = (x_1(t), \cdots, x_n(t))$, is given by

$$\frac{dR}{dt} = \frac{\partial R}{\partial c^i}\frac{dc^i}{dt} = \frac{dc^i}{dt}\boldsymbol{e}_i$$

where $\boldsymbol{e}_i$ are the basis vectors for coordinate system (c^i). So this is a basis expansion of the derivative, and we get that component of dR/dt are dc^i/dt.

So what are the transformation rules for dR/dt? Is it contravariant or covariant? Let us expand it into two coordinate systems $(c^i), (p^j)$:

$$\begin{aligned}\frac{dR}{dt} &= \frac{\partial R}{\partial c^i}\frac{dc^i}{dt} && \text{by chain rule}\\ &= \frac{\partial R}{\partial p^j}\frac{dp^j}{dc^i}\frac{dc^i}{dt} && \text{by backward transformation rule}\\ &= \frac{\partial R}{\partial p^j}\frac{dp^j}{dt} && \text{by direct expansion in coordinates } p^j\end{aligned}$$

So we have a transformation rule for components:

$$\frac{dp^j}{dt} = \frac{dp^j}{dc^i}\frac{dc^i}{dt}$$

we used backward transformation to go from new coordinates (p^j) to old coordinates (c^i). Therefore, components of the derivative are contravariant.

From now on, we will only write $\boldsymbol{e}_i = \frac{\partial}{\partial x}$ because the choice of R does not matter. Another main reason is that for curved surfaces, $\frac{\partial R}{\partial x}$ is not straight. So instead of talking about vectors, we use the partial derivative operators. Since derivative operators are linear, they still form a linear space.

3.2 Differentials are tensors

Recall that dx is the infinite small change in the x direction, and it is used for integration.

4 Appendices

4.1 Hyperbolic functions

Recall the hyperbolic functions are defined as

$$\sinh x = \frac{e^x - e^{-x}}{2}, \quad \cosh x = \frac{e^x + e^{-x}}{2}$$

And other hyperbolic functions are defined using $\sinh, \cosh$

$$\tanh x = \frac{\sinh x}{\cosh x},\ \coth x = \frac{\cosh x}{\sinh x},\ \operatorname{sech} x = \frac{1}{\cosh x},\ \operatorname{csch} x = \frac{1}{\sinh x}$$

The hyperbolics are related to trig functions via

$$\sinh x = -i \sin ix,\ \cosh x = \cos ix$$

all properties of hyperbolic can be derived from this, but I will still list them here.

Identities:

$$\cosh^2 x - \sinh^2 x = 1,\ \operatorname{sech}{}^2 x = 1 - \tanh^2 x,\ \operatorname{csch}{}^2 x = \coth^2 x - 1$$

Angle addition formulae:

$$\sinh(x+y) = \sinh x \cosh y + \cosh x \sinh y$$

$$\cosh(x+y) = \cosh x \cosh y + \sinh x \sinh y$$

$$\tanh(x+y) = \frac{\tanh x + \tanh y}{1 + \tanh x \tanh y}$$

These formulae are very similar to those for trig functions, but some signs are different so please be careful.