General Relativity Marathon Notes

Aiman Al-Eryani

November 21, 2018

Contents

1 Events, Inertial Frames of Reference and Galilean Relativity **2**
1.1 Example of an inertial frames of reference . . . 2
1.2 Non-example of inertial frames of reference . . . 2
1.3 Galilean transformations . . . 2
1.4 Three reference frames . . . 3

2 Special Relativity **5**
2.1 Motivation . . . 5
2.2 Postulates of Special Relativity . . . 5
2.3 The new transformation . . . 5
2.4 Finding the Lorentz transformation . . . 6
2.4.1 Finding $\alpha(u)$ in terms of $\gamma(u)$ and showing that $\gamma(u) = \gamma(-u)$: . . . 6
2.4.2 Finding $\theta(u)$ and $\tau(u)$ in terms of $\gamma(u)$ . . . 7
2.4.3 Finding the function $\gamma(u)$ . . . 8
2.5 Sorting the mess, space-interval contraction and time-interval dialation . . . 8
2.6 Lorentz transformations in three dimensions . . . 10
2.7 Lorentz covariance . . . 11
2.8 Spacetime diagrams: space and time on equal footing . . . 12
2.8.1 Metrics . . . 13
2.8.2 The minkowski metric . . . 14
2.9 Physical and geometric interpretation and the light cone . . . 15
2.9.1 Example: The twin paradox . . . 18
2.10 Algebra of Special Relativity . . . 18
2.10.1 Groups . . . 18
2.10.2 The Lorentz and the Poincare Group . . . 19

3 General Relativity **19**
3.1 Motivation . . . 19
3.2 The principle of general covariance . . . 20
3.3 The Einstein equivalence principle . . . 21
3.3.1 The weak equivalence principle . . . 21
3.3.2 The equivalence principle . . . 21
3.4 Tensors and coordinate basis . . . 21
3.4.1 Vectors as rank-1 tensors . . . 21
3.4.2 Matrices as rank-2 tensors . . . 23
3.4.3 Higher rank tensors . . . 25

3.5 Transformation rules: contravariant VS covariant and an alternative definition of a tensor . . . 25
3.5.1 Einstein notation . . . 25
3.5.2 Change of basis transformation, and a new definition of a tensor . . . 27
3.5.3 The dual of a vector and index gymnastics . . . 29
3.6 The Geodesic Equation from the einstein equivalence principle . . . 31

1 Events, Inertial Frames of Reference and Galilean Relativity

- An *event* is defined a point $\vec{x}$ in space at a certain time t. We write them together $(t, \vec{x})$
- An *inertial frame of reference* is a non-accelerating frame of reference. Alternatively, it's a reference frame where Newton's laws holds. In particular, it's usually Newton's second law that gets violated in non-inertial frames of reference (acceleration appears without a force).

1.1 Example of an inertial frames of reference

- If you set up a lab on an airplane, Newton's second law will still hold; moving objects keep moving without accelerating and stationary objects remain stationary unless an external force intervenes (friction, pushing ..etc). So the frame of the lab is an inertial frame of reference. You can place a cup of water on your table and even play a game of pool with your friend.
- If you set up a lab on the surface of the earth (neglecting that the earth rotates), Newton's second law will also apply.

1.2 Non-example of inertial frames of reference

- If you set up your small lab on a roller coaster, then you attempt to place your pool balls stationary on the table, your game will quickly be interrupted everytime the roller coaster takes a turn or go through a loop. You'll also quickly find that cup of water has also moved and that the water is no longer in the cup. Stationary objects don't remain stationary despite you not applying any forces on them. This violates Newton's second law, and so this frame is not inertial.

1.3 Galilean transformations

- For the two inertial frames of reference: the earth's and the plane's. Is the plane's lab moving and the earth is stationary, or is the earth moving and the plane is stationary?
- If you have your axes such that the origin is always moving with the lab on the plane, you're doing physics *in the reference frame of the plane*
- Similarly, if you have your axes such that the they're always moving in the center of your lab on earth, then you're doing physics *in the reference frame of the earth*
- There's **no absolute inertial frame** or one that is special; all inertial frames are just as good.
- Suppose a ball is stationary in the plane's frame (with respect to the origin on the plane's lab) at position x_P and time t_P. What would be the event of the ball (t'_P, x'_P) at a later time t'_P?

$$\begin{aligned} x'_P &= x_P && \text{;ball is stationary, so new position equals old position} \\ t'_P &= t'_P && \text{;the new time is the new time} \end{aligned}$$

- What about in the reference frame of the earth? What would be the event of the ball at a later time t'_E if the plane is moving away from the lab on earth at velocity u?

$$\begin{aligned} x'_E &= x'_P + ut'_P && \text{;position evolves as ball is moving in this frame} \\ t'_E &= t'_P && \text{;time on the plane equals the time on earth later} \end{aligned}$$

- If the plane is stationary with respect to the lab on earth, $u = 0$ and:

$$\begin{aligned} x'_E &= x'_P + ut'_P = x'_P \\ t'_E &= t'_P \end{aligned}$$

which tells us that the two frames are exactly the same.

- These transformations from one reference frame to another are called *Galilean transformations*.
- We can see that the relative velocity between the two reference frames is what differentiates them and is what determines what the galilean transformation.
- We'll denote a galilean transformation from frame A to another frame B moving with velocity u relative to A as G_u.
- Using fancy vector notation, we could have written the transformations above from frame of the plane P to frame of the earth E with G_u being a matrix and the event as a vector:

$$\begin{pmatrix} t'_E \\ x'_E \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ u & 1 \end{pmatrix} \begin{pmatrix} t'_P \\ x'_P \end{pmatrix} = \begin{pmatrix} t'_E \\ ut'_P + x'_P \end{pmatrix}$$

so G_u can be expressed as a matrix:

$$G_u := \begin{pmatrix} 1 & 0 \\ u & 1 \end{pmatrix}$$

which you multiply the event vector by.

1.4 Three reference frames

- Suppose we have three inertial frames of reference. One reference frame T attached to a tree, another reference frame C attached to a canoe (or a boat), and one more reference frame B attached to a ball.
- The ball - and therefore the frame B attached to it - is moving with velocity v relative to the canoe - or to its frame C.
- The canoe with its frame C is moving with velocity u with repsect to the frame tree with its frame T.

- The ball is always stationary in the reference frame B since it's attached to it. Therefore in B, the event of the ball at a later time t'_B is:

$$x'_B = x_B$$
$$t'_B = t'_B$$

- What about in the reference frame of the canoe C? We know the velocity of the ball in this frame is v, so we just use a galilean transformation to get the evolved coordinates in the frame C, just like before:

$$x'_C = x'_B + ut'_B$$
$$t'_C = t'_B$$

- What about in the reference frame of the tree T? We only know that the canoe is moving with velocity u relative to the tree, but we don't know yet how the ball is moving relative to the tree (you do know it'll be $u + v$, but that's what we're trying to prove). What would be (t'_T, x'_T)?

- Idea: We do a two-way galilean transformation

$$(t'_B, x'_B) \xrightarrow{G_v} (t'_C, x'_C) \xrightarrow{G_u} (t'_T, x'_T)$$

Let's compute that manually, and use the fancy vector notation:

$$\begin{aligned}
\text{We know that:} \quad \begin{pmatrix} t'_C \\ x'_C \end{pmatrix} &= G_v \begin{pmatrix} t'_B \\ x'_B \end{pmatrix} \\
\text{and that:} \quad \begin{pmatrix} t'_T \\ x'_T \end{pmatrix} &= G_u \begin{pmatrix} t'_C \\ x'_C \end{pmatrix} \\
\text{substituting:} \quad \Longrightarrow \begin{pmatrix} t'_T \\ x'_T \end{pmatrix} &= G_u G_v \begin{pmatrix} t'_B \\ x'_B \end{pmatrix} \\
&= \begin{pmatrix} 1 & 0 \\ u & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ v & 1 \end{pmatrix} \begin{pmatrix} t'_B \\ x'_B \end{pmatrix} \\
\begin{pmatrix} t'_T \\ x'_T \end{pmatrix} &= \begin{pmatrix} 1 & 0 \\ u + v & 1 \end{pmatrix} \begin{pmatrix} t'_B \\ x'_B \end{pmatrix} = \begin{pmatrix} t'_B \\ (u + v)t'_B + x'_B \end{pmatrix}
\end{aligned}$$

The main result here is that applying G_v to go from B to C then G_u to go from C to T is equivalent to applying G_{u+v} to go immediately from B to T. This tells us that *velocities add up*. This is called the **Galilean law addition of velocities**. It's something that we all know very well; that if you can pitch a ball at 100 mph, then the catcher would recieve it at 100 mph. But if you do it while on top of a skateboard moving at 10 mph, then the catcher will recieve a 100 mph + 10 mph = 110 mph fast ball.

However this will turn out not to be the case in great generality. In particular, with speeds near the speed of light, velocities don't add up like that.

2 Special Relativity

2.1 Motivation

- What if you shine light in reference frame B, and then view it in frame A that moving with velocity u w.r.t frame B? Galilean law of addition of velocities tells you that the velocity of light would be $v + c$.
- Problem: Maxwell's theory of electromagnetism suggests that light is an electromagnetic wave, and that all electromagnetic waves travel at speed $c = \frac{1}{\sqrt{\mu_0 \varepsilon_0}}$.
- Both the magnetic constant μ_0 and the electric constant ε_0 are constants that have the same value in all reference frames, and c should there just be $\frac{1}{\sqrt{\mu_0 \varepsilon_0}}$ in all frames. Contradiction?
- People concluded that light, like any wave, must be travelling in a medium that they called 'ether'. Ether can be dragged around with a reference frame so Maxwell's theory still holds.
- Michelson-Morley would then in 1887 do an experiment involving mirrors and making use of interference properties and show that, contrary to what they would have liked, that light does not travel in a medium.
- So either Maxwell or Newton and his company did some mistake. Einstein sided with Maxwell.

2.2 Postulates of Special Relativity

Einstein set himself two postulates which his new theory should follow:

1. The laws of physics are the same in all inertial frames of reference.
2. The speed of light c is the same in all inertial frames of reference.

We can see that both postulates were taken one step further from what we've taken previously. We previously only wanted Newton's laws to be the same in all inertial frames, but now it's all the laws of physics. In the second postulate, Einstein did not only postulate that *electromagnetic waves* in particular travel in c in all reference frames; rather he took the speed c as special, and anything that travels at c must travel at c in all inertial frames.

2.3 The new transformation

- Before we start thinking of a new transformation that would keep c constant in all inertial frames, we must consider that for $\frac{v}{c} << 1$, this transformation must correspond to the galilean transformation. This is known as *the correspondence principle.*
- The plot of the position of object x against time t is called its *spacetime diagram.*
- It's clear that the spacetime diagram of an object moving at velocity u is a straight line, as long as there are no forces acting on it. This is just Newton's second law, as any accleration will result in a curve.
- For the Newton's second law to hold, it must mean that going from one reference frame to another should use a transformation that keeps the straight lines in spacetime diagrams straight, since any curve implies acceleration (change in slope is a change in velocity).

- Geometrically, a galilean transformation 'shears' (sort of "stretches") the spacetime diagram in the direction of x. Shearing can never make straight lines in spacetime diagrams curve.

- We also know that the new transformation we're looking for must also keep the straight lines straight for Newton's second law to hold in all reference frames. A change between inertial frames shouldn't introduce any acceleration.

- In other words, the transformation is linear. Any linear transformation can be represented by a matrix.

- Let's call denote this transformation matrix by L_u, where u is a velocity characterising it. The most general form of this matrix is:

$$L_u = \begin{pmatrix} \theta(u) & \tau(u) \\ \alpha(u) & \gamma(u) \end{pmatrix}$$

where $\theta(u)$, $\tau(u)$, $\alpha(u)$ and $\gamma(u)$ are unknown functions of u. What we'll do next is find these functions, and therefore determine the matrix L_u.

2.4 Finding the Lorentz transformation

- Letting go of any previously held notions of galilean transformations of how velocities add up, the following are things we least expect from this new transformation L_u:

$$L_u \begin{pmatrix} t \\ -ut \end{pmatrix} \overset{!}{=} \begin{pmatrix} t' \\ 0 \end{pmatrix}$$; in a frame where your frame is moving at u, an object that is moving at $-u$ in your frame should look stationary

$$L_{-u}L_u = L_u L_{-u} \overset{!}{=} I$$; this should be identity for $u \neq c$

$$L_u \begin{pmatrix} t \\ ct \end{pmatrix} \overset{!}{=} \begin{pmatrix} t' \\ ct' \end{pmatrix}$$; the second postulate: the speed of light c is the same in any frame

$$L_0 \overset{!}{=} I$$; going from a frame to another comoving frame must be identity

$$L_c L_u \neq L_{c+u}$$; velocities should not add up in the galilean sense

- Only the first three conditions are needed to find L_u completely. The last two will just be used to test that we have arrived at the correct result.

- Note that with galilean transformations, time never transformed; it would always be the same no matter what reference frame we pick. However, here, we do not exclude that the time component gets transformed. And indeed, as we shall see, t is not always equal to t'.

2.4.1 Finding $\alpha(u)$ in terms of $\gamma(u)$ and showing that $\gamma(u) = \gamma(-u)$:

- Starting from the first condition:

$$\begin{pmatrix} \theta(u) & \tau(u) \\ \alpha(u) & \gamma(u) \end{pmatrix} \begin{pmatrix} t \\ -ut \end{pmatrix} = \begin{pmatrix} t\theta(u) - ut\tau(u) \\ t\alpha(u) - ut\gamma(u) \end{pmatrix} p \overset{!}{=} \begin{pmatrix} t' \\ 0 \end{pmatrix}$$

$$t\alpha(u) - ut\gamma(u) = 0 \implies \boxed{\alpha(u) = u\gamma(u)}$$

- Let's update the matrix L_u so that it no longer has $\alpha(u)$:

$$L_u = \begin{pmatrix} \theta(u) & \tau(u) \\ u\gamma(u) & \gamma(u) \end{pmatrix}$$

- The second condition is that the inverse of all of our transformations L_u must be L_{-u}. Let's try to hit the expression of the first condition with L_{-u} on both sides to see if we can squeeze out any more info from it:

$$\cancel{(L_{-u})L_u} \begin{pmatrix} t \\ -ut \end{pmatrix} \overset{!}{=} (L_{-u}) \begin{pmatrix} t' \\ 0 \end{pmatrix} = \begin{pmatrix} \theta(-u) & \tau(-u) \\ -u\gamma(-u) & \gamma(-u) \end{pmatrix} \begin{pmatrix} t' \\ 0 \end{pmatrix}$$
$$\begin{pmatrix} t \\ -ut \end{pmatrix} \overset{!}{=} \begin{pmatrix} t'\theta(-u) \\ -ut'\gamma(-u) \end{pmatrix}$$

we have $-ut = -ut'\gamma(-u) \implies t = t'\gamma(-u)$. By the isotropy of space (time should transform the same way regardless of what direction u is), we should also have that $t = t'\gamma(u)$. This implies that $\boxed{\gamma(u) = \gamma(-u)}$. The same argument can be repeated with the equation between elements of the first components, and we get that $\theta(-u) = \theta(u)$.

2.4.2 Finding $\theta(u)$ and $\tau(u)$ in terms of $\gamma(u)$

- Now let's see what we can get if we expand the expression of the second condition:

$$L_{-u}L_u \overset{!}{=} I$$
$$\begin{pmatrix} \theta(-u)\theta(u) + \tau(-u)(u\gamma(u)) & \theta(-u)\tau(u) + \tau(-u)\gamma(u) \\ (-u\gamma(-u))\theta(u) + \gamma(-u)(u\gamma(u)) & (-u\gamma(-u))\tau(u) + \gamma(-u)\gamma(u) \end{pmatrix} \overset{!}{=} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

looks very scary but its just ordinary matrix multiplication. These are four equalities. Let's look at the equality in the left-bottom corner of the matrix:

$$(-u\gamma(-u))\theta(u) + \gamma(-u)(u\gamma(u)) \overset{!}{=} 0$$

Recall that $\gamma(u) = \gamma(-u)$. We can then solve for $\theta(u)$ in terms of $\gamma(u)$:

$$-u\gamma(u)\theta(u) + u\gamma^2(u) \overset{!}{=} 0 \implies \boxed{\theta(u) = \gamma(u)}$$

- Now let's look at the bottom-right corner and try to solve for $\tau(u)$. We find that:

$$(-u\gamma(-u))\tau(u) + \gamma(-u)\gamma(u) \overset{!}{=} 1 \implies -u\gamma(u)\tau(u) + \gamma^2(u) = 1$$
$$\implies \boxed{\tau(u) = \frac{\gamma^2(u) - 1}{u\gamma(u)}}$$

- We can now express the entire matrix in terms of $\gamma(u)$!

$$L_u = \begin{pmatrix} \gamma(u) & \frac{\gamma^2(u)-1}{u\gamma(u)} \\ u\gamma(u) & \gamma(u) \end{pmatrix}$$

now all is left to determine L_u is to determine the function $\gamma(u)$.

2.4.3 Finding the function $\gamma(u)$

- Taking the third condition, which is essentially invoking the second postulate of special relativity (that an object moving at the speed of light should do so in all reference frames):

$$L_u \begin{pmatrix} t \\ ct \end{pmatrix} \stackrel{!}{=} \begin{pmatrix} t' \\ ct' \end{pmatrix}$$

$$\begin{pmatrix} \gamma(u) & \frac{\gamma^2(u)-1}{u\gamma(u)} \\ u\gamma(u) & \gamma(u) \end{pmatrix} \begin{pmatrix} t \\ ct \end{pmatrix} \stackrel{!}{=} \begin{pmatrix} t' \\ ct' \end{pmatrix}$$

$$\begin{pmatrix} t\gamma(u) + ct\frac{\gamma^2(u)-1}{u\gamma(u)} \\ tu\gamma(u) + ct\gamma(u) \end{pmatrix} \stackrel{!}{=} \begin{pmatrix} t' \\ ct' \end{pmatrix}$$

we have two equalities and two unknowns $\gamma(u)$ and t'. Either plug them into Worlfram alpha or solve them. Plugging the expression for t' (the first equality) in the second equality:

$$\cancel{t}u\gamma(u) + \cancel{ct\gamma(u)} = \cancel{c(t\gamma(u))} + c(c\cancel{t}\frac{\gamma^2(u)-1}{u\gamma(u)})$$

$$\implies \boxed{\gamma(u) = \frac{1}{\sqrt{1-\frac{u^2}{c^2}}}}$$

we have thus fully determined the new transformation matrix L_u.

- Let's keep using the variable $\gamma(u)$, but let's try to simplify the term in the upper right corner of L_u: $\frac{\gamma^2(u)-1}{u\gamma(u)} = (\frac{1}{u} - \frac{1}{u\gamma^2(u)})\gamma(u) = (\frac{1}{u} - \frac{1-\frac{u^2}{c^2}}{u})\gamma(u) = \gamma(u)(\cancel{\frac{1}{u}} - \cancel{\frac{1}{u}} + \frac{u}{c^2}) = \frac{u}{c^2}\gamma(u)$
- We finally get:

$$\boxed{L_u = \begin{pmatrix} \gamma(u) & \frac{u}{c^2}\gamma(u) \\ u\gamma(u) & \gamma(u) \end{pmatrix}}$$

- The function $\gamma(u)$ is called the *Lorentz factor*, and the transformation L_u is called the *Lorentz transformation*.
- **Excercise:** Check that the last two conditions do indeed hold, and find the new formula for addition of velocities.
- **Excercise:** The fact that c is the same in all reference frames implies that an object that's not already moving at c can not be moving at c in any reference frame. Check that this is indeed the case by trying to compute L_c.

2.5 Sorting the mess, space-interval contraction and time-interval dialation

- **First, we need to sort out one mess**. We have been talking about L_u as the transformation that takes you from your frame to another frame in which according to it you're moving at velocity u. This was easier to calculate, but not so easy to phrase. In other literature, you'll see the parameter of the transform u being the velocity with which that frame is moving relative to your frame. In other words, they define $\Lambda_u = L_{-u}$. We'll switch to that convention from now on.

- We explicitly write Λ_u:

$$L_{-u} = \boxed{\Lambda_u = \begin{pmatrix} \gamma(u) & -\frac{u}{c^2}\gamma(u) \\ -u\gamma(u) & \gamma(u) \end{pmatrix}}$$

- Let's see how arbitrary events (t, x) transform under the lorentz transformation Λ_u that takes us to frame S' that's moving at u:

$$\Lambda_u \begin{pmatrix} t \\ x \end{pmatrix} = \begin{pmatrix} \gamma(u) & -\frac{u}{c^2}\gamma(u) \\ -u\gamma(u) & \gamma(u) \end{pmatrix} \begin{pmatrix} t \\ x \end{pmatrix} = \begin{pmatrix} \gamma(u)(t - \frac{u}{c^2}x) \\ \gamma(u)(x - ut) \end{pmatrix} \stackrel{!}{=} \begin{pmatrix} t' \\ x' \end{pmatrix}$$

- Suppose you have in your frame two events both at the same time t but one at position x and the other at position $x + \Delta x$. One can think the distance between them is:

$$\begin{pmatrix} t \\ x + \Delta x \end{pmatrix} - \begin{pmatrix} t \\ x \end{pmatrix} = \begin{pmatrix} 0 \\ \Delta x \end{pmatrix}$$

- What would be the distance in the reference frame S'? We need to transform both events and then compute the difference again. By linearity of matrices, it's the same as taking the difference and then computing:

$$\Lambda_u \begin{pmatrix} t \\ x + \Delta x \end{pmatrix} - \Lambda_u \begin{pmatrix} t \\ x \end{pmatrix} = \Lambda_u(\begin{pmatrix} t \\ x + \Delta x \end{pmatrix} - \begin{pmatrix} t \\ x \end{pmatrix}) = \Lambda_u \begin{pmatrix} 0 \\ \Delta x \end{pmatrix}$$
$$\implies (\Delta x)' = \begin{pmatrix} -\frac{u\Delta x}{c^2} \\ (\Delta x)\gamma(u) \end{pmatrix}$$

we have a 'spacital distance' $(\Delta x)\gamma$, but it doesn't help because the events are no longer simultaneous (!). We have a distance in space as well a distance in time in the new frame! Taking the spacial distance wouldn't make sense because it would be the difference between two events at two different times.

- What does that tell us? Two events that are simultaneous in one inerital frame may not be in the other. So "what happened first, a chicken or an egg?" might have different answers in different reference frames.

- In general, there is a trade-off between length contraction and simultaneouity. In most special relativity "paradoxes", the trick is to remember that.

- Now that we know time differences can change, how does the time interval between two events at the same position $(\Delta t + t, x)$ and (t, x) look like in the a different frame S'? In the current frame, the difference is just $(\Delta t, 0)$. Let's Lorentz transform it:

$$\Lambda_u \begin{pmatrix} \Delta t \\ 0 \end{pmatrix} = \begin{pmatrix} \Delta t\gamma(u) \\ -u\Delta t\gamma(u) \end{pmatrix} \stackrel{!}{=} \begin{pmatrix} \Delta t' \\ \Delta x' \end{pmatrix}$$

we see that here the spatial distance changes, but that does not matter for time measurement. Therefore, $\Delta t' = \Delta t\gamma(u)$. We have that $\gamma(u) \geq 1$, which means that time dialates when moving; or as they say, *moving clocks tick faster*. This effect is known as *time dialation.*

- A clock at rest is one constructed from events which x component has no velocity term ut for any velocity u involved. If a clock at rest measures the (proper) time interval $\Delta\tau$, then in another moving frame, it would measure $\Delta t = \tau\gamma(u)$ as we have seen above due to time dialation. This implies $\boxed{\Delta\tau = \frac{\Delta t}{\gamma(u)}}$. This is an invariant quantity that's the same in all frames.

- But if time is changing in every frame, what would be the right time to measure? It's simply the time where the clock is at rest. By definition, it's the same in all reference frames. We call it the *proper time* and we define it by $\Delta\tau$ (or sometimes τ).

- With a similar idea, we can define a proper length L in frame that is at rest. One can show in a similar way to proper time that it's $L_0 = \gamma(u)L$ once they have found length contraction.

- **Excercise:** Show that *length contraction*; show that the $\Delta x' = \frac{1}{\gamma(u)}\Delta x$ where $\Delta x'$. Avoid simultaneity issues we encounted before by ensuring $\Delta t' = 0$ in the new frame S'. Hint: Start with Δt arbitrary and then insist that $\Delta t' = 0$. After that, solve for Δt and substitute in the equation for $\Delta x'$.

2.6 Lorentz transformations in three dimensions

- So far, we have only worked with x position and time, but the world has three spatial dimensions.

- Claim: if motion is strictly on the x axis (no velocity term for y and z), transformations for y and z are trivial: $y = y'$, $z = z'$.

- Proof: Since Lorentz transformations are linear, they must map $y = 0$ to $y' = 0$, and $z = 0$ to $z' = 0$. If $y, z \neq 0$, we just translate everything so that they're at zero. Changing the origin of the coordinate system should not change its physics (in general: changing the coordinate system shouldn't affect the physics. That's the principle of *general covariance*).

- So we have:

$$\Lambda_u \begin{pmatrix} t \\ x \\ y \\ z \end{pmatrix} = \begin{pmatrix} \gamma(u) & -\frac{u}{c^2}\gamma(u) & 0 & 0 \\ -u\gamma(u) & \gamma(u) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} t \\ x \\ y \\ z \end{pmatrix} = \begin{pmatrix} \gamma(u)(t - \frac{u}{c^2}x) \\ \gamma(u)(x - ut) \\ y \\ z \end{pmatrix} \stackrel{!}{=} \begin{pmatrix} t' \\ x' \\ y' \\ z' \end{pmatrix}$$

- What if u is not strictly in the x direction? We rotate everything so that u is in the x-axis direction with the rotation matrix R^{-1}, do our physics (lorentz transform), and then rotate back with R. This is equivalent to just hitting L_u by "conjugation" of R (means, hit on the right with inverse, and on the left with itself):

$$\Lambda_{R,u} = R\Lambda_u R^{-1}$$

and then use the new $\Lambda_{R,u}$.

2.7 Lorentz covariance

- In physics, we deal with quantities. But we need good ones that are well defined and consistent. It's a bit vague at the moment, but soon it should be made clear.

- Scalars quantities like the spatial distance between two events Δx takes different values in different inertial frames (due to length contraction), and is therefore bad. A good quantity is one that's *Lorentz invariant*; a quantity that is the same in all inerital frames. For example, the proper time $\Delta\tau$ is Lorentz invariant. It's the same in all reference frames. Other examples of Lorentz invariant quantities are the rest mass m_0, the speed of light c, the proper length and so forth.

- Any meaningful vector quantities which involve the physical dimensions Length or Time will have to transform under a lorentz transformation. We base that of the fact that an event that describes a position-time in spacetime transforms by application of the Lorentz transform operator L_u. Velocity for instance is then the derivative of this position, and momentum would be a mass times this position ..etc.

- However, this one disturbing thing with having an event as a quantity of position. A quantity should have one physical dimension (Length, Time ..etc). However, the first component in the event vector has dimension Time, and the rest have dimension Distance/Spactial-Position.

- To make everything consistent, we must make the first components of an event also have a dimension of Spacial-Position. We know the way to make a distance from a time quantity is to multiply it by a velocity quantity.

- In order for the event to still transform similarly, the velocity we multiply by time better be an invariant, and we know that got to be c. This way, it will simply act as a scaling factor.

- We define the *4-position*:

$$X = \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix}$$

our base position quantity in special relativity. To account for the scaling in time, the Lorentz transformation changes form:

$$\Lambda = \begin{pmatrix} \gamma(u) & -\frac{u}{c}\gamma(u) & 0 & 0 \\ -\frac{u}{c}\gamma(u) & \gamma(u) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

one can easily check that applying this on the the 4-Position vector X gives the same equations when an event without a scaling factor c is hit by the old Lorentz transform.

- One can try to define a velocity out of this. From the formal definition of a derivative, $\frac{dX}{dt} = \lim_{\Delta t \to 0} \frac{\Delta X}{\Delta t}$. The difference between two 4-vectors will transform by multiplication by Λ under a change of inertial frames:

$$
\begin{aligned}
&\Delta X = X_1 - X_2 \\
\overset{\text{go to frame } S'}{\rightarrow} \; &\Delta X' = X_1' - X_2' = \Lambda X_1 - \Lambda X_2 = \Lambda(X_1 - X_2) = \Lambda \Delta X \\
\overset{\text{derivative limit}}{\rightarrow} \; &dX' = \Lambda dX
\end{aligned}
$$

However, Δt is a scalar that's not a lorentz scalar, so will change. Instead, we use the proper time $\Delta\tau$:

$$V = \frac{dX}{d\tau}$$

- Since $d\tau$ is invariant, V transforms like dX which transforms like X: with multiplication by Λ.
- We can express $d\tau$ as follows:

$$\Delta\tau = \frac{1}{\gamma}\Delta t \implies d\tau = \frac{1}{\gamma}dt$$

$$\implies \boxed{V = \gamma\frac{dX}{dt} = \gamma\begin{pmatrix} c \\ \dot{x} \\ \dot{y} \\ \dot{z} \end{pmatrix}}$$

this is called the *4-velocity*.

- Similarly, we can define other quantities like 4-momentum as $P = m_0 V = m_0\gamma\frac{dX}{dt}$ where m_0 is the rest mass, a Lorentz invariant. And then the 4-force $\frac{dP}{d\tau} = \gamma\frac{dP}{dt}$, which is also a four vector. And the list goes on. These are all quantities that respect Lorentz transformation.
- In general, if all quantities in a law respect lorentz transformations, then the law is lorentz covariant and is compatable with special relativity.
- Maxwell equations, in particular, are lorentz invariant; SR was based to make Maxwell's equations hold afterall. We know how the velocity of light transforms for example; it's a lorentz invariant. One can also check that other quantities form nice four vectors and everything. More will be taught in the second year course "Electrodynamics".
- Newton's law fo gravitation however, does not respect lorentz transofrmations. In particular, it has explicit dependence on the distance between two bodies r, which is not a lorentz invariant nor a lorentz covariant.

2.8 Spacetime diagrams: space and time on equal footing

- So we now have events described by (ct, x, y, z). All components have physical dimension of length. Thus, we can say that an event is a quantity now also known as the *4-position*, the spacetime position of an object.
- Furthermore, 4-position is a lorentz covariant; it transforms "correctly" and consistently under a change of the inertial frame of reference (transforms by multiplication by Λ).
- We saw that after scaling the time component of an event by c to define the $4-position$, the Lorentz transformation becomes (ignoring the y and z components for simplicity):

$$\Lambda = \begin{pmatrix} \gamma(u) & -\frac{u}{c}\gamma(u) \\ -\frac{u}{c}\gamma(u) & \gamma(u) \end{pmatrix}$$

which looks slightly nicer, as it has now became symmetric if the movement is only on one spacial axis x. Suppose we take the event vector and swap the time axis with the event axis so that it's (x, ct) instead of (ct, x). The lorentz transformation matrix would not be exactly the same!

- This tells us that space and time are to be treated on equal footing. This was not the case in Galilean transformation:

$$G = \begin{pmatrix} 1 & 0 \\ u & 1 \end{pmatrix}$$
$$G^T = \begin{pmatrix} 1 & u \\ 0 & 1 \end{pmatrix}$$

where the matrix is not symmetric and changing (t, x) to (x, t) would require we take G^T.

- If we were to talk about "Galilean invariants", there are plenty. The speed of light c is not a galilean invariant (light speed is different in different inertial frames). Δx is an invariant in galilean relativity; changing inertial frames doesn't change the distance between two points. To generalize to 3D, the invariant space interval Δs is the galilean sense is given by:

$$\Delta s = \sqrt{\Delta x^2 + \Delta y^2 + \Delta z^2} \xrightarrow{\text{infinitisimal}} ds^2 = dx^2 + dy^2 + dz^2$$

- With lorentz', Δx and dx is not an invariant. Going to a boosted frame (a frame that's moving) causes length contraction. But it also causes time dialation. One wonders whether this trade off makes a combination of both a lorentz invariance (spoiler: yes).

2.8.1 Metrics

- There's a name given to an object that takes two objects and gives the distance between them. This object is called the *metric* and can be denoted by $m(a, b) = ||a - b||$. It takes two points $\vec{a}$ and $\vec{b}$ and gives a scalar that's the distance between them. The metric above can be written as:

$$m(\vec{a}, \vec{b}) = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + (a_3 - b_3)^2}$$

- Another way to define the distance between two points $\vec{\Delta x}$ is by taking the square root of its dotproduct with itself since:

$$\vec{\Delta x} \cdot \vec{\Delta x} = ||\vec{\Delta x}||^2 = \Delta x_1^2 + \Delta x_2^2 + \Delta x_3^2$$

Metrics defined that way are called "induced metrics". In physics, the dotproduct is so important it's often called the metric, even though the value of the distance is defined by its square root.

- So from now on, when we say the "metric", we mean the "inner product". Afterall, what we care most about is whether its value is invariant or not. If the distance is invariant, then so should be its square. When we want distance, we'll say "the distance" or "the interval" explicitly.

- The dot product/metric has two very important properties. The first is that it's symmetric; namely that:

$$\vec{a}\cdot\vec{b}=\vec{b}\cdot\vec{a}$$

The second is that it's bilinear; You can split sums and take out/in factors in its two argument "slots":

$$(\alpha\vec{a}+\beta\vec{b})\cdot\vec{c}=\alpha(\vec{a}\cdot\vec{c})+\beta(\vec{b}\cdot\vec{c})$$

Since it's symmetric, one can see that if it's linear in one of the arguments, then it's also linear in the other (since we can swap the arguments).

- The generalisation of linear operators with any number of slots that give back numbers are called *tensors*.
- Matrices are *rank 2 tensors* because they can act on two vectors and give a number. So given a matrix A, you can evaluate it on vectors $\vec{a}$, $\vec{b}$ to get a number by taking the first vector transposed, multiplying on the left, and multiplying the other vector on the right: $\vec{a^T}A\vec{b}$.
- Using a matrix, we can write the dotproduct/metric above as:

$$\vec{a}\cdot\vec{b}=\begin{pmatrix}a_1 & a_2 & a_3\end{pmatrix}\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix}\begin{pmatrix}b_1\\b_2\\b_3\end{pmatrix}=a^TIb=a_1b_1+a_2b_2+a_3b_3$$

where the matrix in the middle is the identity matrix. So we see that the euclidean metric is really the identity matrix.

2.8.2 The minkowski metric

- Now here is what we're interested in: is there a notion of "space-time" distance that's invariant under lorentz transformations?
- Another way to say it: Is there a metric which if we evaluate on an ΔX we get a lorentz invariant?
- Another way to say it: Is there a metric which the lorentz transformation is an *isometry* of? (an isometry is a map that preserves the distances/evaluations of the metric).
- Let's try to find this metric η. We would use to find the "spacetime" distance between two points δs by applying the metric (our "dot product") on the difference vector of two events $\Delta X=X_1-X_2$ and taking the square root:

$$\Delta s=\sqrt{(\Delta X)^T\eta(\Delta X)}\implies(\Delta s)^2=(\Delta X)^T\eta(\Delta X)$$

- Going to a different inertial frame, the difference vector would transform as $\Delta X\to\Lambda\Delta X$. Yet, the spacetime distance should stay the same. This means:

$$
\begin{aligned}
(\Delta s)^2 = (\Delta X)^T \eta (\Delta X) \stackrel{!}{=} (\Lambda \Delta X)^T \eta (\Lambda \Delta X) = (\Delta X^T) \Lambda^T \eta \Lambda (\Delta X) \\
\iff \eta \stackrel{!}{=} \Lambda^T \eta \Lambda
\end{aligned}
$$

- After staring at the equation for a while, one can find that:

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

- η is called the **Minkowski metric**, and it's the metric of spacetime.
- Let's try to compute the distance between X_2 and X_1 denoted by $\Delta X = X_1 - X_2$:

$$
\begin{aligned}
\Delta X^T \eta \Delta X &= \begin{pmatrix} c\Delta t & \Delta x & \Delta y & \Delta z \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} c\Delta t \\ \Delta x \\ \Delta y \\ \Delta z \end{pmatrix} \\
&= -(c\Delta t)^2 + (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2
\end{aligned}
$$

- Rotations are transformations that keep the euclidean distance the same. Gometrically, it's movement on locus of a circle.
- In this sense, a lorentz transformation is one that keeps the minkowski distance invariant. Geometrically, it's movement on locus of a hyperbola.
- Things to notice: We can apply this to any 4-vector and that Δs^2 can go negative! In mathematics lingo, this makes it a "pseudo-metric". However we shall continue to refer to it by simply "metric".
- **Excercise:** Show that the lorentz invariant magnnitude of the 4-velocity is c ! It means we're moving in c all the time, and that thus $\Delta \tau^2 = -\Delta s^2$

2.9 Physical and geometric interpretation and the light cone

- If one draws a spacetime diagram of ct in the y-axis vs x in the x-axis, one can see that the trajectories $x(t) = ct$ and $x(t) = -ct$ are diagonal lines. These define a *light cone* or the event horizon.
- Any flatter slope would imply a trajectory of something moving faster than c which is not possible. So the boundary of the light cone defines an event horizon outside which nothing can influence anything inside it or viceversa.
- Spacial rotation matrices are the isometries of the euclidean metric. On an x-y diagram, a rotation operation will make the vector $\Delta X = (\Delta x, \Delta y)$ trace contour lines of $\Delta s^2 = \Delta x^2 + \Delta y^2$ with Δs being fixed; that is the locus of a circle of radius Δs.

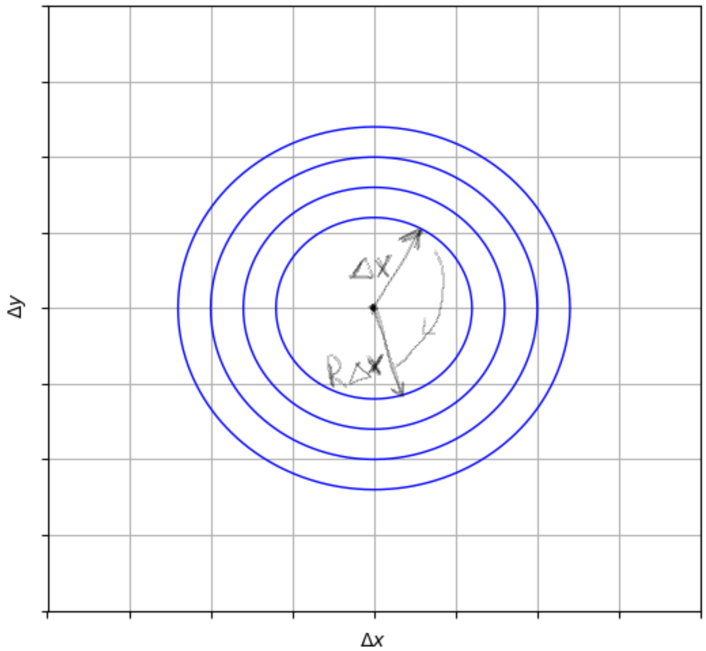

Figure 1: Multiplying ΔX by a rotation matrix R, an isometry of the euclidean metric, is geometrically just sliding the vector on the locus of a circle of fixed radius

- Analogously, Lorentz transformation matrices (boosts) are the isometries of the minkowski metric. On a spacetime diagram with ct against x, a lorentz transformation matrix will make the vector $\Delta X = (c\Delta t, \Delta x)$ trace the contour lines of $\Delta s^2 = -(c\Delta t)^2 + \Delta x^2$. This equation can refer to different shapes depending on the value of Δs^2:
  - If $\Delta s^2 < 0 \implies \Delta x^2 = \pm\sqrt{\pm(c\Delta t)^2 + |\Delta s^2|}$, it will trace the locus of a hyperbola either concaving upwards or downwards. Events with $\Delta s < 0$ are called **timelike separated**. From the geometric picture, it's clear one can always perform a lorentz transformation that makes ΔX point fully up (or fully down); i.e. one can always use a lorentz transformation two make two time-like separated events have zero spatial separation. However, no matter what, you can never "slide" the vector on a locus to make it horizontal (i.e. zero time-sparation or simultaneous) with $\Delta t = 0$.
  - If $\Delta s^2 > 0 \implies \Delta x^2 = \pm\sqrt{\pm(c\Delta t)^2 + |\Delta s^2|}$, it will trace the locus of a hyperbola either concaving leftwards or rightwards.Events with $\Delta s < 0$ are called **timelike separated**. From the geometric picture, it's clear one can always perform a lorentz transformation that makes ΔX point fully up (or fully down); i.e. one can always use a lorentz transformation two make two space-like separated events have zero time separation or be simultaneous! However since this event is outside the light cone, one can never observe this. Conversely to above, you can never "slide" the vector on a locus to make it vertical (i.e. zero space-sparation) with $\Delta x = 0$.
  - If $\Delta s^2 = 0 \implies c\Delta t = \pm\Delta x$, it will trace the locus of a line with slope ± 1. Events with $\Delta s^2 = 0$ are called **null separated**; which means one of the two events is moving at the speed of light. From the geometric picture, since the lines start at the origin, the vector ΔX is always fully intersecting with the line and thus there's nowhere new it can

move; no lorentz transformation can do anything to change it. This is consistent with the second postulate of Special Relativity which states that the speed of light is the same in all inertial reference frames.

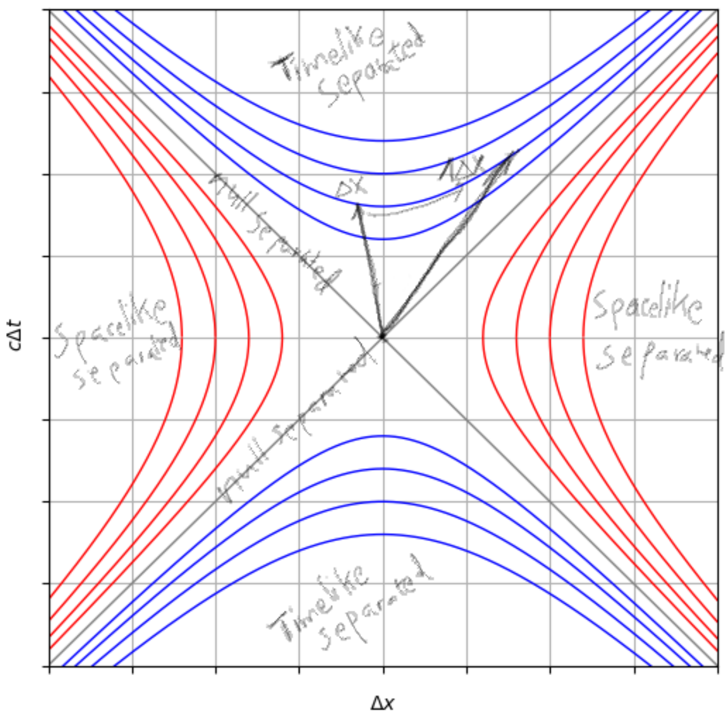

Figure 2: Multiplying the difference between two events ΔX by a lorentz transformation matrix Λ, an isometry of the minkowski metric, is geometrically just sliding the vector on the locus of a hyperbola

- Timelike separated events, i.e. these for which ΔX is inside the light cone, are causally connected and can influence one another. Spacelike separated events with their ΔX outside the light cone are causally disconnected and can never influence one another.
- Suppose you're in a frame where you're at rest. You measure the time Δt, and by definition, since you're at rest, the time you measured is the proper time $\Delta\tau$. So in this case $\Delta t = \Delta\tau$. Note that since you're at rest, the spacial distance you moved is $\Delta x = 0$. The spacetime distance is then

$$\Delta s^2 = -(c\Delta t)^2 + \cancel{(\Delta x)^2} = -(c\Delta t)^2 \implies \boxed{\Delta s^2 = -(c\Delta\tau)^2}$$

- The left handside of the boxed equation is a spacetime distance which is the same in all reference frames by definition (thus a lorentz invariant). The right handside is a product of two lorentz invariants, namely the speed of light c and the proper time $\Delta\tau$. This means that the equation above holds regardless of the inertial frames you're at, and regardless of the two events you're measuring distances between. It gives an expression for the propertime $\Delta\tau$.
- Notice the proper time will only give you a real number between timelike separated events, and is thus only physical for timelike separated events.

2.9.1 Example: The twin paradox

- **Problem:** Alice and Bob each start measuring time at the same point. Bob travels away from Alice at velocity v. After covering a distance d, he makes a sudden turn and returns back to Alice. What would be the time measured for both Alice and Bob?

- **Solution:**

  - Since Bob makes a turn somewhere, the frame attached to him while he took his full trajectory is not inertial. So we should stick to Alice's inertial frame of reference.
  - Since Alice is stationary in her frame, $\Delta x_A = 0$.
  - From Alice's perspective, it takes Bob $\Delta t = \frac{2d}{v}$ seconds to return, so the spacetime interval between the start of the measurement and its end is $\Delta s_A^2 = -(c\Delta t)^2 = -4d^2\frac{c^2}{v^2}$.
  - Bob, however, will have in addition traveled a distance of $\Delta x = v\Delta t = 2d$. So his spacetime interval would be $(\Delta s_B)^2 = -4d^2\frac{c^2}{v^2} + 4d^2$.
  - Now the time each measured is their proper time for both. We have that $(\Delta s)^2 = -(c\Delta\tau)^2 \implies \Delta\tau = \sqrt{\frac{-(\Delta s)^2}{c^2}}$

$$\begin{aligned}
\text{So for Alice: } \Delta\tau_A &= \sqrt{\frac{-(\Delta s_A)^2}{c^2}} = \sqrt{\frac{4d^2\frac{c^2}{v^2}}{c^2}} \\
&= \frac{2d}{v} \\
\text{For Bob: } \Delta\tau_B &= \sqrt{\frac{-(\Delta s_B)^2}{c^2}} = \sqrt{\frac{4d^2\frac{c^2}{v^2} - 4d^2}{c^2}} \\
&= \frac{2d}{v}\sqrt{1 - \frac{v^2}{c^2}}
\end{aligned}$$

  - So we have that $\Delta\tau_A = \Delta\tau_B\sqrt{1 - \frac{v^2}{c^2}}$ which tells us that $\Delta\tau_A > \Delta\tau_B$ (more time passed for Alice than Bob).

2.10 Algebra of Special Relativity

2.10.1 Groups

- Formally a group is a set with an operation on its elements $K(x, y)$ that satisfies four axioms:

  1. Has an "identity element" I such that $K(x, I) = x$ and $K(I, x) = x$
  2. Each element has an "inverse element" x^{-1} such that $K(x, x^{-1}) = I$ and $K(x^{-1}, x) = I$
  3. Assocativity: $K(K(x, y), z) = K(x, K(y, z))$
  4. Closedness: The operation K always produces an element inside our group.-

- Example: Group of all points in $\mathbb{R}^3$, and let the operator be the addition of the two position vectors (i.e. translation).

- Example: Group of all linear operators that respect the Euclidean metric (preserve pythagorean distances) in the three dimensional space. These are all the invertible matrices A which inverses are themselves transposed; i.e. $A^{-1} = A^T \implies AA^T = A^TA = I$. Its fancy name is the *Special Orthogonal Group* or $SO(3)$. "Special" means invertible, and "Orthogonal" means $A^T = A^{-1}$. These matrices are rotation matrices.

2.10.2 The Lorentz and the Poincare Group

- Group of all matrices which respect the Minkowski metric (preserves the spacetime distance). These are the matrices that obey $A\eta A^T = g$ where η is the minkowski metric.
- We have seen before that Lorentz transformation matrices and matrices that rotate in space obey this relation (in fact, that's how we found η in the first place).
- This group of spacial rotation and boosts (Lorentz transformations) is called *the Lorentz group* and is denoted by $SO(3, 1)$.
- One can thing of it as the group of rotations on spacetime, where rotation on spacetime is on a locus of a hyperbola rather than a circle.
- The Lorentz group with the group of operators that translate on spacetime is called *the Poincare group* and is denoted by $TSO(3, 1)$.
- Note that $TSO(3, 1)$ is not a linear group, however all its elements are isometries of the Minkowski metric.

3 General Relativity

"Spacetime tells matter how to move. Matter tells spacetime how to curve. "
- John Wheeler

3.1 Motivation

- Newton's law of gravitation says that the magnitude of the force acting on the earth from the the sun is:

$$F_G = \frac{GM_{sun}m_{earth}}{(x_{earth} - x_{sun})^2}$$

  If the sun suddenly disappears, it tells us that since $M_{sun} = 0$ the earth immediately stops feeling any force and leaves orbit. But this is in contradition with special relativity that tells us that it would stop feeling the force at least after ~8 minutes have passed.
- In particular, Newton's law of gravitation is not Lorentz covariant. The three dimensional spacial quantity $x_{earth} - x_{sun}$ is not a four vector nor an invariant.

- Another motivation was the descrepency between Mercury's orbit and the predictions Newton's gravity makes. Newton's gravity predicts orbits follow elliptical paths; the orbits never cross. However in the case of Mercury, the orbit crosses itself and so the perihelion of the ellipse of ist orbit precesses. Therefore a new theory was needed.

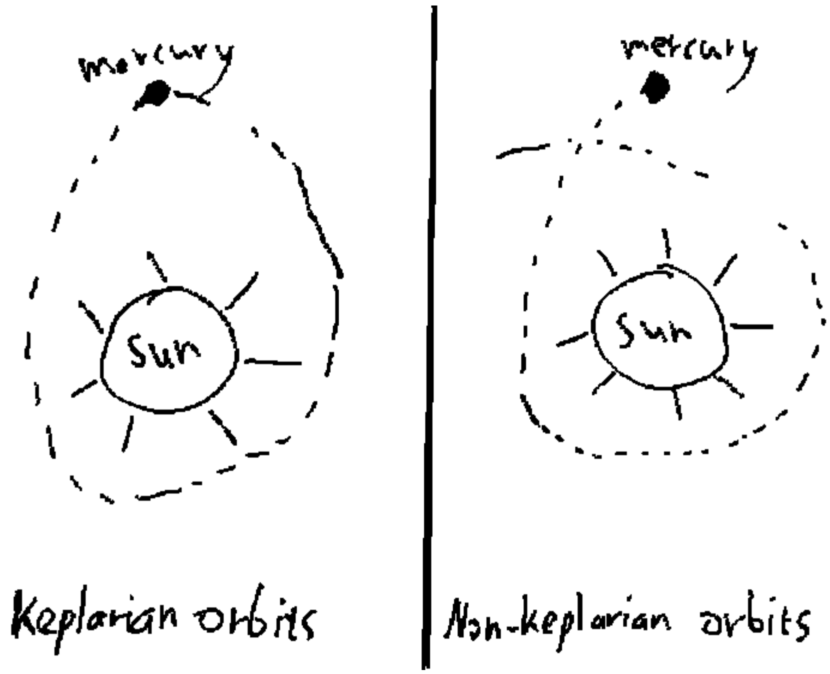

Figure 3: On the left is a keplarian orbit where the planet follows an ellipse. On the right is a non-keplarian orbit that's observed with Mercury

- Einstein's idea: gravity is geometric and gravitional acceleration is just freefalling. Gravitional acceleration is just an emergent property of curved spacetime.
- Special Relativity is a special case of General Relativity where the *spacetime is flat*.
- In General Relativity, spacetime curves in response to the presence of matter. Acceleration then emerges naturally from curvature of spacetime.
- Main concept: Curved space can be covered by multiple patches that are approximately flat. The structure that contains all these patches is called a *manifold*.
- At each point, there's a different flat metric; or one says there's a *metric field* defined on each point on the manifold. The spacetime distance between two events would then be the sum (integral) of all the little distances on the patches you pass by.
- A manifold equipped with a metric field is called a *Riemanian Manifold*. In the case of general relatvitiy, the metric at each point will be pseudoriemanian (we'll show they all can take negative distances like the Minkowski metric), so we will be working with *Pseudo-Riemanian manifolds*. The maths of (Pseudo)Riemanian manifolds is "Riemanian Geometry" which we will have to cover to quantify the ideas of General Relativity.

3.2 The principle of general covariance

- We have understood that lorentz transformations are just (hyperbolic) rotations in the time-axis.

- The lorentz group ($SO(3,1)$, group of rotations in space and time) is really a group of symmetries where the laws of physics do not change. A law that's invariant under its transformations is said to be lorentz covariant.
- More generally, translations is also a symmetry, and so the poincare group ($TSO(3,1)$) sets a similar condition for *poincare covariance.*
- In general, these are all really just coordinate transformations. We expect the laws of physics to be the same under *all* coordinate transformations. That's the principle of **General Covariance**.
- We shall see that the principle of general covariance greatly restricts the form of physics laws. We will thus focus on studying objects that obey the priniciple of general covariance, and then we will see later that once some preeliminary choices have been made, there are very few ways to write the equations of general relativity.

3.3 The Einstein equivalence principle

3.3.1 The weak equivalence principle

- Inertial mass and gravitational mass are equivalent; i.e. the inertial mass in $F = ma$ and the gravitataional mass in $F = GMm/r^2$ are one and same. This idea is just experimental; there just doesn't seem to be any descripancies, and it seems we can always cancel them out with one another. This was the main motivation behind the idea that gravity must be geometric.

3.3.2 The equivalence principle

- Assume building a little laboratory inside an elevator that's accelerating down uniformly. You take a tiny ball, and let it freefall. The acceleration is indestinguishable from that of gravity.
- But if you take another ball and put it far away, with a high resolution measurement, you may be able to see that the balls are attracting falling into the same center of mass of the earth.
- What you can do: decrease distance between the balls until the measuremnt device can no longer detect it. You can always do this for any resolution of the device.
- So we say **locally** (in a small neighbohrhood), uniform acceleration can't be destinguished from gravity.
- If curvature is responsible for gravity, that's equivalent to saying that locally spacetime is flat in some coordinate system (can involve frames not necessarily inertial).
- In other words, there's a coordinate transformation which makes the metric at point a be $g(a) = \eta$ i.e. a frame that's inertial where acceleration is zero.

3.4 Tensors and coordinate basis

3.4.1 Vectors as rank-1 tensors

- A vector can be represented as a linear combination of basis vectors. E.g: given a basis

$$\mathcal{B} = \{ \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \}$$

, one can write any four dimensional vector A with components A^0, A^1, A^2, A^3 as a sum:

$$\vec{A} = \begin{pmatrix} A^0 \\ A^1 \\ A^2 \\ A^3 \end{pmatrix} = A^0 \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + A^1 \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} + A^2 \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} + A^3 \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

$$\vec{X} = \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix} = ct \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + x \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} + y \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} + z \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

- The choice of basis as above with a 1 in one coordinate and 0s in the others is called the canonical basis. However they're not any special. The basis vectors here could have been any other four linearly independent vectors. One needs as many basis vectors as there are dimensions.
- So for four dimensions, given a basis $\{\hat{e}_0, \hat{e}_1, \hat{e}_2, \hat{e}_3\}$, $\vec{A} = \sum_{\mu=0}^{3} A^\mu \hat{e}_\mu$.
- One can think of row vectors, for example $\vec{B} = \begin{pmatrix} B_1 & B_2 & B_3 & B_4 \end{pmatrix}$. One can also give row vectors a basis, but it would consist of row vectors instead of column vectors. For example:

  $$\mathcal{D} = \{ \begin{pmatrix} 1 & 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 & 1 \end{pmatrix} \}$$

  , or in general $\{\hat{e}^0, \hat{e}^1, \hat{e}^2, \hat{e}^3\}$. And then $\vec{B} = \sum_{\mu=0}^{3} B_\mu \hat{e}^\mu$
- These are called the dual vectors or covectors, because they act on vectors to give a number. We have defined that

  $$\vec{B}\vec{A} = \begin{pmatrix} B_1 & B_2 & B_3 & B_4 \end{pmatrix} \begin{pmatrix} A^0 \\ A^1 \\ A^2 \\ A^3 \end{pmatrix} = B_0 A^0 + B_1 A^1 + B_2 A^2 + B_3 A^3$$

  .
- One should get the same result if they used $\vec{B} = \sum_{\mu=0}^{3} B_\mu \hat{e}^\mu$ and $\vec{A} = \sum_{\nu=0}^{3} A^\nu \hat{e}_\nu$:

  $$\begin{aligned} \vec{B}\vec{A} = (\sum_{\mu=0}^{3} B_\mu \hat{e}^\mu)(\sum_{\nu=0}^{3} A^\nu \hat{e}_\nu) =& B_0 A^0 \hat{e}^0 \hat{e}_0 + B_0 A^1 \hat{e}^0 \hat{e}_1 + B_0 A^2 \hat{e}^0 \hat{e}_2 + B_0 A^3 \hat{e}^0 \hat{e}_3 + \\ & B_1 A^0 \hat{e}^1 \hat{e}_0 + B_1 A^1 \hat{e}^1 \hat{e}_1 + B_1 A^2 \hat{e}^1 \hat{e}_2 + B_1 A^3 \hat{e}^1 \hat{e}_3 + \\ & B_2 A^0 \hat{e}^2 \hat{e}_0 + B_2 A^1 \hat{e}^2 \hat{e}_1 + B_2 A^2 \hat{e}^2 \hat{e}_2 + B_2 A^3 \hat{e}^2 \hat{e}_3 + \\ & B_3 A^0 \hat{e}^0 \hat{e}_0 + B_3 A^1 \hat{e}^0 \hat{e}_1 + B_3 A^2 \hat{e}^3 \hat{e}_2 + B_3 A^3 \hat{e}^3 \hat{e}_3 \end{aligned}$$

  we see that it would be equal to $B_0 A^0 + B_1 A^1 + B_2 A^2 + B_3 A^3$ if and only if *the relation between the vector basis and the dual basis is*:

  $$\hat{e}^i \hat{e}_j = \delta^i_j = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$$

- A dual vector thus operates on a vector to give a real number. But one can also think of it from a different view; the vector operates on the dual vector to give a real number.
- The set of all vectors defined as above is called a vector space (in the case of four dimensional vectors with real number components, it's denoted by $\mathbb{R}^4$, whereas the set of all dual vectors defined as above is denoted is called the dual vector space (for four dimensional real number components, it's denoted by $\mathbb{R}^{4*}$). Basis vectors of a vector space will be denoted by lower indices $\hat{e}_i$, whereas basis vectors of dual vector spaces will be denoted by upper indices $\hat{e}^i$.

3.4.2 Matrices as rank-2 tensors

- To simplify examples here, we'll work with two dimension. A matrix as we have seen is a rank-2 tensor since it can feeds on two vectors to give out a real number.
- Take for instance the following 2x2 matrix with four components

$$A = \begin{pmatrix} A^1{}_1 & A^1{}_2 \\ A^2{}_1 & A^2{}_2 \end{pmatrix}$$

. A basis for 2x2 matrices consists of four matrices:

$$\{\hat{e}_1{}^1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \hat{e}_1{}^2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \hat{e}_2{}^1 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \hat{e}_2{}^2 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}\}$$

, and any two by two matrix can then be represented as a linear combination of these:

$$A = A^1{}_1 \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + A^1{}_2 \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} + A^2{}_1 \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} + A^2{}_2 \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = \sum_{\nu=1}^{2}\sum_{\mu=1}^{2} A^\mu{}_\nu \hat{e}_\mu{}^\nu$$

- A rank 2 tensor can feed on one vector to give another vector (in matrix representation, multiplying the matrix by a vector gives a row vector). So given a vector $\vec{v} \in \mathbb{R}^2$:

$$A\vec{v} = \begin{pmatrix} A^1{}_1 & A^1{}_2 \\ A^2{}_1 & A^2{}_2 \end{pmatrix} \begin{pmatrix} v^1 \\ v^2 \end{pmatrix} = \begin{pmatrix} A^1{}_1 v^1 + A^1{}_2 v^2 \\ A^2{}_1 v^1 + A^2{}_2 v^2 \end{pmatrix} = (A^1{}_1 v^1 + A^1{}_2 v^2)\hat{e}_1 + (A^2{}_1 v^1 + A^2{}_2 v^2)\hat{e}_2$$
$$= A^1{}_1 v^1 \hat{e}_1 + A^1{}_2 v^2 \hat{e}_1 + A^2{}_1 v^1 \hat{e}_2 + A^2{}_2 v^2 \hat{e}_2 = \sum_{\nu=1}^{2}\sum_{\mu=1}^{2} A^\mu{}_\nu v^\nu \hat{e}_\mu$$

- One should get the same result if they used:

$$A\vec{v} = (\sum_{\nu=1}^{2}\sum_{\mu=1}^{2} A^\mu{}_\nu \hat{e}_\mu{}^\nu)(\sum_{\alpha=1}^{2} v^\alpha \hat{e}_\alpha) = A^1{}_1 v^1 \hat{e}_1{}^1 \hat{e}_1 + A^1{}_1 v^2 \hat{e}_1{}^1 \hat{e}_2 + A^1{}_2 v^1 \hat{e}_1{}^2 \hat{e}_1 + A^1{}_2 v^2 \hat{e}_1{}^2 \hat{e}_2$$
$$+ A^2{}_1 v^1 \hat{e}_2{}^1 \hat{e}_1 + A^2{}_1 v^2 \hat{e}_2{}^1 \hat{e}_2 + A^2{}_2 v^1 \hat{e}_2{}^2 \hat{e}_1 + A^2{}_2 v^2 \hat{e}_2{}^2 \hat{e}_2$$

- One can see that the product would match the result of the matrix multiplication $\sum_{\nu=1}^{2}\sum_{\mu=1}^{2} A^\mu{}_\nu v^\nu \hat{e}_\mu$ if and only if *the relation between the basis is*:

$$\hat{e}_j{}^i \hat{e}_\alpha = \delta^i_\alpha \hat{e}_j = \begin{cases} 0 & \text{if } i \neq \alpha \\ \hat{e}_j & \text{if } i = \alpha \end{cases}$$

The basis matrix ${\hat{e}_j}^i$ eats a basis vector $\hat{e}_\alpha$ to give a dual basis vector $\hat{e}_j$ only if its second index i matches with the basis vector's index α. If it doesn't match, it's zero. It's equivalent to evaluating $(\hat{e}^i\hat{e}_\alpha)\hat{e}_j = (\delta^i_\alpha)\hat{e}_j$ which adopts the evaluation of the dual basis on the vector basis vector that we found before.

- Similarly, if ones takes a row vector/dual vector $\vec{w} \in \mathbb{R}^2*$, and computes $\vec{w}A$:

$$\vec{w}A = \begin{pmatrix} w_1 & w_2 \end{pmatrix} \begin{pmatrix} {A^1}_1 & {A^1}_2 \\ {A^2}_1 & {A^2}_2 \end{pmatrix} = \begin{pmatrix} w_1{A^1}_1 + w_2{A^2}_1 & w_1{A^1}_2 + w_2{A^2}_2 \end{pmatrix} = \sum_{\nu=1}^{2}\sum_{\mu=1}^{2} w_\mu {A^\mu}_\nu \hat{e}^\nu$$

  Then by comparing the above $\sum_{\nu=1}^2 \sum_{\mu=1}^2 v_\mu {A^\mu}_\nu \hat{e}^\nu$ as we have done before, one can see the basis need to satisfy:

$$\hat{e}^\beta {\hat{e}_j}^i = \hat{e}^i \delta^\beta_j = \begin{cases} 0 & \text{if } \beta \neq j \\ \hat{e}^i & \text{if } \beta = j \end{cases}$$

  Now the left index of the matrix basis vector j collapses with the dual basis vector's index β leaving a vector basis $\hat{e}^i$ when $\beta = j$. It's equivalent to evaluating $(\hat{e}^{beta}\hat{e}_j)\hat{e}_i = (\delta^\beta_j)\hat{e}_i$ which adopts the evaluation of the dual basis vector on the basis vector on its other index.

- This tells one that ${\hat{e}_i}^j$ action on vectors and dual vectors is really defined in terms of the actions of the dual vectors on vectors and vice versa. This means that the matrix basis vectors are some mix of dual basis vectors and vector basis vectors that feeds on other vectors or dual vectors (respectively) individually. We may thus write it as follows:

$${\hat{e}_j}^i = \hat{e}_j \otimes \hat{e}^i$$

  where $\otimes$ is the *tensor product*. This is the statement that a basis of a rank-2 tensor is just the tensor product of the corresponding basis of the two rank-1 tensors it acts on.

- We thus have the following relations between the bases of the rank-2 tensors and rank-1 tensors:

$$\begin{aligned} &\hat{e}_j \otimes \hat{e}^i(\hat{e}_\alpha) = \hat{e}_j \cdot (\delta^i_\alpha) && \text{Matrix basis multiplied by a basis vector from the right} \\ &(\hat{e}^\alpha)\hat{e}_j \otimes \hat{e}^i = (\delta^\alpha_j) \cdot \hat{e}^i && \text{Matrix basis multiplied by a dual basis vector from the left} \end{aligned}$$

- We say that this tensor is then an element of a *tensor product space* of the two vector spaces $\mathrm{V} \otimes \mathrm{V}^*$; in this case, $\mathbb{R}^2 \otimes \mathbb{R}^{2*}$. It has two slots and is linear in both. In the first slot, it feeds on a dual vector (row vector in the matrices representation), and in the second it feeds on a (column) vector to finally give a real number.

- We have been a bit sloppy by saying it's a rank 2 tensor since it feeds on two vector, but it actually feeds on 1 vector and one dual vector, so it's classified as rank $(1, 1)$ tensor. What we usually mean when we say its rank is two is that its total rank is two. We'll see later however that this slopiness is justified for our purposes, since we will fix a transformation (an isomorphism) between vectors and their duals. So we will easily be able to transform a rank (p, q) tensor into a rank $(p + q, 0)$ tensor. So we just call both a rank $p + q$ tensor.

- We see that the dimension $\dim(\mathrm{V} \otimes \mathrm{V}^*) = \dim \mathrm{V} \times \dim \mathrm{V}^* = (\dim \mathrm{V})^2$.

3.4.3 Higher rank tensors

- A rank (p,q) tensor is an element of a tensor product space $\overbrace{V^* \otimes \cdots V^*}^{\text{p-times}} \otimes \overbrace{V \otimes \cdots V}^{\text{q-times}}$, which feeds on p vectors and q dual vectors to give out a number.

- Examples:
  •Any scalar is a rank $(0,0)$ tensor.

  • $v(\overbrace{\ \_\ }^{\in V^*}) \in V$ is a rank $(1,0)$ tensor (or a rank 1 tensor).

  It has one slot for a dual vector, so for $w \in V^*$, $v(w)$ is a scalar. In matrices representation, we know this evaluation is just multiplication of the vector v by the row vector w from the left; wv.

  • $w(\overbrace{\ \_\ }^{\in V}) \in V^*$ is a rank $(0,1)$ tensor (or a rank 1 tensor).

  It has one slot for a dual vector, so for $v \in V$, $w(v)$ is a scalar. In matrices representation, we know this is also wv. So we see that a dual vector acting a vector $w(v)$ is the same as the vector acting on the dual $v(w)$.

  • $A(\overbrace{\ \_\ }^{\in V^*}, \overbrace{\ \_\ }^{\in V}) \in V \otimes V^*$ is a rank 2 tensor

  Filling the first slot $A(v, \_)$ with $v \in V$ makes it a rank $(0,1)$ tensor (a dual vector). In matrix representation, it's equivalent to multiplying the column vector v from the right; the result of Av is indeed a column vector. Filling the second slot $A(\_, w)$ with $w \in V^*$ makes it a rank $(1,0)$ tensor (a vector). In matrix represention, it's multiplying the row vector w from the left of the matrix; wA. Filling both slots $A(v,w)$ give a rank $(0,0)$ tensor which is a scalar. In matrix representation, it's equivalent to multiplying the matrix from the left and the right wAv, and this indeed gives a scalar.

- Alternatively, one can fill two slots with one rank-2 tensor. Or three slots with one rank-1 tensor and one rank-2 tensor, and so forth.

- In general relativity, we will be dealing with tensors of higher ranks (in particular, the Riemann curvature tensor is a rank-4 tensor). For tensors of total rank higher than two, matrices representation no longer help (even if you have an idea of how to still use them, perhaps with nesting matrices within matrices ..etc, it become very inconvenient). This motivates Einstein's summation convention.

3.5 Transformation rules: contravariant VS covariant and an alternative definition of a tensor

3.5.1 Einstein notation

- Assume we are working on a vector space V of dimension $n < \infty$ (its dual vector space V^* will always have the same dimension). Assume furthermore that the field over which the vectorspace is defined is $\mathbb{R}$ (i.e. that's a fancy way of saying the components of the tensors are real numbers). Here are some tensors:

- $k \in \mathbb{R}$ is a rank (0, 0) tensor (a scalar).
- $v \in \mathrm{V}$ is a rank (1, 0) tensor (a vector).
- $w \in \mathrm{V}^*$ is a rank (0, 1) tensor (a dual vector).
- $M \in \mathrm{V} \otimes \mathrm{V}^*$ is a rank (1, 1) tensor.
- $T \in \mathrm{V} \otimes \mathrm{V}^* \otimes \mathrm{V}$ is a mixed tensor of total rank 3 (sort of a rank $(2, 1)$ tensor, but not quite since the order of the slots matters).

- Einstein's idea:

1. **Write the tensors in terms of a sum of their components multiplied by basis vectors. Then ignore writing the sums and the basis vectors**:

$$\begin{aligned}
&k && \xrightarrow{\text{Einstein's notation}} && k \\
&v = \textstyle\sum_{i=1}^{n} v^i \hat{e}_i && \xrightarrow{\text{Einstein's notation}} && v^i \\
&w = \textstyle\sum_{i=1}^{n} w_i \hat{e}^i && \xrightarrow{\text{Einstein's notation}} && w_i \\
&M = \textstyle\sum_{i=1}^{n} \sum_{j=1}^{n} M^i{}_j \hat{e}_i \otimes \hat{e}^j && \xrightarrow{\text{Einstein's notation}} && M^i{}_j \\
&T = \textstyle\sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} T^i{}_j{}^k \hat{e}_i \otimes \hat{e}^j \otimes \hat{e}_k && \xrightarrow{\text{Einstein's notation}} && T^i{}_j{}^k
\end{aligned}$$

From just looking at the symbol with the indices, one can deduce the rank of the tensor. For instance, take $M^i{}_j$:

- It has two indices in total, so it has total rank 2.
- It has 1 upper index, and 1 lower index, so it feeds on a dual vector and a vector to give a scalar; it has rank $(1, 1)$.

Similarly, $T^i{}_j{}^k$ has total rank 3, and has three slots in the order: Vector, Dual Vector, Vector. v^i has one upper index, so it feeds on a dual vector to give a scalar; i.e. it's a vector or a rank $(1, 0)$ tensor. One can in addition express the tensor product of two tensors by just sticking the tensors together with different indices. For instance, $v^i w_j$ is collectively a rank-2 tensor, and $M^i{}_j v^k w_l$ is a rank-4 tensor.

2. **Repeated indices imply a sum over the indices**:

$$w(\underline{v}) = (\sum_{i=1}^{n} w_i \hat{e}^i)(\sum_{j=1}^{n} v^j \hat{e}_j) = \sum_{i=0}^{n}\sum_{j=0}^{n} w_i v^j \overbrace{\hat{e}^i(\hat{e}_j)}^{\delta^i_j} = \overbrace{\sum_{i=0}^{n} w_i v^i}^{\in\mathbb{R}} \xrightarrow{\text{Einstein's notation}} w_i v^i$$

$$M(w, \overbrace{\_\_}^{\in \mathrm{V}}) = (\sum_{i=1}^{n}\sum_{j=1}^{n} M^i{}_j \hat{e}_i \otimes \hat{e}^j)(\sum_{k=1}^{n} w_k \hat{e}^k, \_\_)$$

$$= \sum_{i=0}^{n}\sum_{j=0}^{n}\sum_{k=0}^{n} M^i{}_j w_k \overbrace{(\hat{e}_i(\hat{e}^k)) \otimes \hat{e}^j}^{\delta^k_i \cdot \hat{e}^j} = \sum_{i=0}^{n}\sum_{j=0}^{n} M^i{}_j w_i \hat{e}^j \xrightarrow{\text{Einstein's notation}} M^i{}_j w_i$$

Note how pairing upper and lower indices by setting them equal is equivalent to filling a slot. Therefore, repeated indices don't count towards the rank. So for example, $M^\mu{}_\mu$ has zero free indices, so it's a scalar. The name given to the operation of pairing two indices this way is called **contraction**. Note that, in general, contracting two indices takes the tensor from rank p to rank $p - 2$.

3.5.2 Change of basis transformation, and a new definition of a tensor

- Given a basis $\mathcal{A} = \{\hat{e}_i\}$, and another basis $\mathcal{A}' = \{\hat{e}_{i'}\}$ for a vector space V, one can transform the $\mathcal{A}$ basis to the primed basis $\mathcal{A}'$ by a linear transformation Q (we'll use matrix representation):

$$\boxed{Q\hat{e}_i = \hat{e}_{i'} \implies \hat{e}_i = Q^{-1}\hat{e}_{i'}} \xrightarrow{\text{Einstein's notation}} \boxed{Q^i{}_{i'}\hat{e}_i = \hat{e}_{i'} \implies \hat{e}_i = Q^{-1i'}{}_i\hat{e}_{i'}}$$

Let's take a vector $\vec{v} \in$ V and write it down as a linear combination of basis vectors: $\vec{v} = \sum_{i=1}^n v^i\hat{e}_i$. To express $\vec{v}$ in terms of the other primed basis, since we know $\hat{e}_i = Q^{-1}\hat{e}_{i'}$, we just substitute. Let's write that in a semi-einstein notation where we don't write the sums but we still write the basis:

$$\vec{v} = v^i\hat{e}_i \xrightarrow{\text{basis change}} v^i(Q^{-1i'}{}_i\hat{e}_{i'}) = (v^iQ^{-1i'}{}_i)\hat{e}_{i'} \xrightarrow{\text{making basis invisible}} \boxed{Q^{-1i'}{}_iv^i =: v^{i'}}$$

We see with the basis transformation, v transforms with the the *inverse* of Q; i.e. **contravariantly**. So the vectors of the vector space $\vec{v} \in$ V are called **contravariant vectors**.

- Now let's look at how the dual basis $\mathcal{D} = \hat{e}^i$ would transform as a result. We have for the old basis that the dual basis and the vector basis satisfy $\hat{e}^i\hat{e}_j = \delta^i_j$. We now require that the new bases also obey the relation:

$$\hat{e}^{i'}\hat{e}_{j'} = \delta^{i'}_{j'} \xrightarrow{\hat{e}_{j'}=Q^j{}_{j'}\hat{e}_j} \hat{e}^{i'}(Q^j{}_{j'}\hat{e}_j) = \delta^{i'}_{j'} \implies \boxed{\hat{e}^{j'} = Q^{-1j'}{}_je^j \implies e^j = Q^j{}_{j'}\hat{e}^{j'}}$$

Let's take a dual vector $\vec{w} \in$ V* and write it as: $\vec{w} = \sum_{i=1}^n w_i\hat{e}^i$. Let's express it in terms of the primed basis while explicitly writing down the basis:

$$\vec{w} = w_i\hat{e}^i = w_i(Q^i{}_{i'}\hat{e}^{i'}) \xrightarrow{\text{making basis invisible}} \boxed{Q^i{}_{i'}w_i =: w_{i'}}$$

We see that upon the change of basis, w transforms *together* with Q; i.e. **co-variantly**. So the vectors of the dual vector space $\vec{w} \in$ V* are called **covariant vectors**.

- What about tensors? Take for example $T \in$ V $\otimes$ V* $\otimes$ V, and let's write that as a linear combination:

$$\begin{aligned} T =& T^{\alpha}{}_{\beta}{}^{\gamma}\hat{e}_\alpha \otimes \hat{e}^\beta \otimes \hat{e}_\gamma \xrightarrow{\text{in primed basis}} T^{\alpha}{}_{\beta}{}^{\gamma}(Q^{-1\alpha'}{}_\alpha\hat{e}_{\alpha'}) \otimes (Q^\beta{}_{\beta'}\hat{e}^{\beta'}) \otimes (Q^{-1\gamma'}{}_\gamma\hat{e}_{\gamma'}) = \\ =& (T^{\alpha}{}_{\beta}{}^{\gamma}Q^{-1\alpha'}{}_\alpha Q^\beta{}_{\beta'}Q^{-1\gamma'}{}_\gamma)\hat{e}_{\alpha'} \otimes \hat{e}^{\beta'} \otimes \hat{e}_{\gamma'} \xrightarrow{\text{rearranging and making basis invisible}} \\ =& \boxed{Q^{-1\alpha'}{}_\alpha Q^\beta{}_{\beta'}Q^{-1\gamma'}{}_\gamma T^{\alpha}{}_{\beta}{}^{\gamma} =: T^{\alpha'}{}_{\beta'}{}^{\gamma'}} \end{aligned}$$

So a *tensor transforms by contracting* Q^{-1} *with each contravariant index and* Q *to each covariant index.*

- In fact, defining a tensor as an *array of numbers that transforms like that* is a perfectly valid and equivalent definition of a tensor. It is actually much more useful in physics to define it like that.

- Let's forget about relativity for a moment, and only consider a system in three dimensional euclidean space (vector space is therefore in $\mathbb{R}^3$). The electric field due to a point charge is given by $\vec{E} = k\frac{Q}{|\vec{r}|^3}\vec{r}$ where k is a constant, Q is the charge of the point source, $\vec{r}$ is the displacement from the point source, $\vec{r}$ is the norm of that (the scalar distance from the point source). We know the object $\vec{r}$ has three components in x, y and z. But is it a covariant vector, a contravariant vector, or not even a vector at all? We can answer this question by seeing how it transforms when we do a change of basis (change of coordinates). Let's choose a new basis where each basis vector gets doubled in size. In this new coordinate system, the components of r' will get reduced by half: it transformed by the inverse of the transformation done on the basis vectors! (which were scaled to twice the size). This means that $\vec{r}$ is a *contravariant* vector, because it transforms like one.

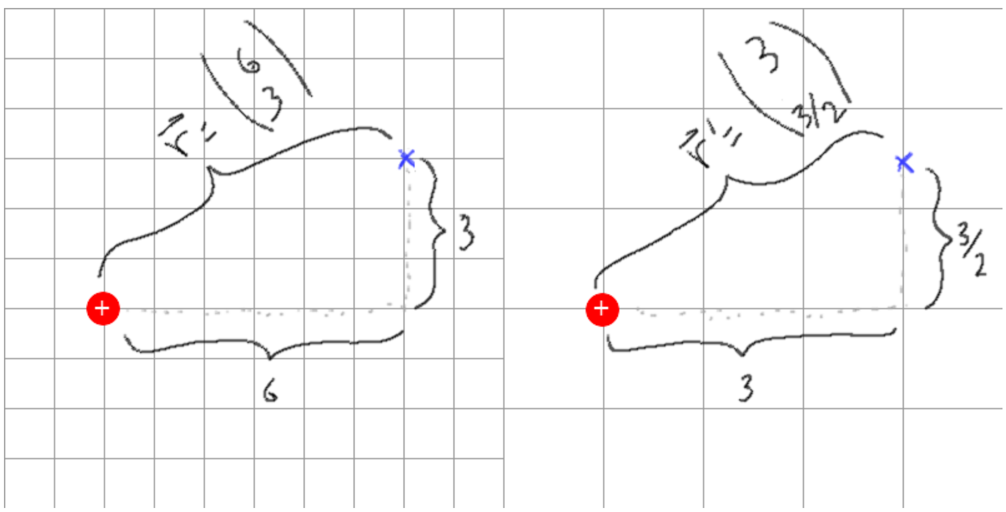

Figure 4: The blue x is the point we're measuing $\vec{E}$ at. The configuration on the left and on the right are identical, but the underlying coordinate system of the figure on the right has been scaled up by 2x

- What about the electric field $\vec{E}$? If the the components of $\vec{r'}$ are half those of $\vec{r}$; $\vec{r'} = \frac{1}{2}\vec{r}$, then by the formula:

$$\vec{E'} = \frac{kQ}{|\vec{r'}|^3}\vec{r'} = \frac{kQ}{\left|\frac{1}{2}\right|^3 |\vec{r}|^3}\frac{1}{2}\vec{r} = 4\frac{kQ}{|\vec{r}|^3}\vec{r} = 4\vec{E}$$

When the basis size was doubled, the components of $\vec{E}$ grew by a factor of four. It transforms covariantly! However it's not a dual vector, since a dual vector should transform by just a factor of two. Recall that transforming together by twice the transformation on the basis is how a *rank* $(0, 2)$ *tensors transform*! That's very very strange. However, if you know a bit more physics that I assumed, then you would know that in special relativity the electric field by itself is not really a dual 4-vector nor a 4-vector (not lorentz covariant), but is actually a component of a rank $(0, 2)$ tensor called the *electromagnetic magnetic field tensor*, which components can be written in matrix representation as follows:

$$F_{\mu\nu} = \begin{pmatrix} 0 & E_x/c & E_y/c & E_z/c \\ -E_x/c & 0 & -B_z & B_y \\ -E_y/c & B_z & 0 & -B_x \\ -E_z/c & -B_y & B_x & 0 \end{pmatrix}$$

This tensor would actually scale up by four under the coordinate transformation above! For more about the covariant formulation of Maxwell's theory, see appendix (##not written yet).

- By thinking of how a bunch of numbers we can observe transform, we can tell what kind of a tensor they are, and this can give us a lot of insight about the kind of geometric object it is. More examples of this include the angular velocity $\vec{\omega}$, which transforms with an extra negative sign if the y-basis vector is flipped. This tells us that it's something called a *differential two form* that we will look at later (in fact, $F_{\mu\nu}$ is also a differential two form. You can check that it transforms with an extra minus sign if only one of the basis vectors sign is flipped.)

3.5.3 The dual of a vector and index gymnastics

- Let's now assume that our vector space is the lorentzian vector space $\mathrm{V} = \mathbb{R}^{3,1}$, where we have four dimensions (three spacial dimensions, and one time dimension). For the indices, it's then customary to start counting from 0 where the 0th component is the time component, and components $1, 2, 3$ are spacial components. So given the 4-position vector X^μ, $X^0 = ct$, $X^1 = x$, $X^2 = y$ and $X^3 = 3$.

- We have seen that contracting a dual vector with a vector gives a scalar. In relativity, we do have scalars that we're interested in; lorentz invariant scalars (like proper time τ, speed of light c, rest mass m_0 and so forth). We have seen that the minkowski metric η operating on any two 4-vectors (say A^μ and B^ν) give a lorentz invariant scalar. By wording it like that, it's a rank $(2, 0)$ tensor. Let's write that in einstein's notation:

$$\eta(\underline{A}, \underline{B}) \xrightarrow{\text{Matrix representation}} A^{\mathrm{T}}\eta B \xrightarrow{\text{Einstein's notation}} A^\mu \eta_{\mu\nu} B^\nu$$

So the metric is a rank $(2, 0)$ tensor, but in matrix representation, it's clearly operating on a column vector and a row vector (dual vector), shouldn't it be a rank $(1, 1)$ tensor?

- Let's first begin by clearing that confusion. The minkowski metric rank that is a rank $(2, 0)$ tensor is actually not represented by the matrix η alone. It's represented by a matrix *together with* 'the transpose operator'. The transpose operator is a linear map; i.e. given $A, B \in \mathrm{V}$ and $c, d \in \mathbb{R}$ then $(cA + dB)^{\mathrm{T}} = c(A^{\mathrm{T}}) + c(B^{\mathrm{T}})$. In this case, it takes a vector and gives back a dual vector (row vector), and a dual vector takes a vector and finally gives a scalar. This means that in total, it takes two vectors and return a scalar; it's actually a rank $(2, 0)$ tensor:

$$v^{\mathrm{T}} = {}^{\mathrm{T}}\overbrace{(v, \overbrace{\_\_}^{\in \mathrm{V}})}^{\in \mathrm{V}^*}$$

So we may only consider one slot and say it's a map from the vector space to the dual space:

$${}^{T}(\overbrace{\_\_}^{\in \mathrm{V}}) \in \mathrm{V}^*$$

- So given the matrix η as a rank $(1,1)$ tensor, one forms the rank $(2,0)$ metric tensor by plugging in the transpose as a map from the vector space to the dual space into one of the slots:

$$\eta(\overbrace{\_}^{\in \mathrm{V}}, \overbrace{\_}^{\in \mathrm{V}^*}) \in \mathrm{V}^* \otimes \mathrm{V} \xrightarrow{\text{Plugging in the transpose to one of the slots}} \eta({}^{\mathrm{T}}(\overbrace{\_}^{\in \mathrm{V}}), \overbrace{\_}^{\in \mathrm{V}}) \in \mathrm{V}^* \otimes \mathrm{V}^*$$

- Now the transpose in the metric's slot over there never really needed to exist. The metric is not a matrix, but a rank $(2,0)$ tensor, and was only used as a matrix because in matrices representation, they're the only objects that have total rank-2. So the tranpose had to be injected to make it give out a scalar. But from now on, the metric η is a rank $(2,0)$ tensor, and is therefore written in einstein's notation with two lower indices as $\eta_{\mu\nu}$.

- It's components are the same (since the transpose doesn't affect components): $\eta_{00} = -1$, $\eta_{11} = \eta_{22} = \eta 33 = 1$, and the rest are 0.

- We're now interested in lorentz invariant scalars. For convenience, we will identify for each 4-vector X^μ a dual vector which will have the same symbol but with a lower index X_μ so that the scalar $X^\mu X_\mu$ is a lorentz invariant. This means:

$$X^\mu X_\mu \stackrel{!}{=} X^\mu X^\nu \eta_{\mu\nu} \implies \boxed{\Delta X_\mu = \Delta X^\nu \eta_{\mu\nu}}$$

  this makes the metric a device that *lowers indices.*

- Since we have identified to each vector a dual vector, this means that we have implicitly assigned to each dual vector a vector with its same symbol; i.e. there should be a way to raise a lowered index. Contracting the metric $\eta_{\mu\nu}$ with an index of a vector $V_\mu = V^\nu \eta_{\mu\nu}$ is in matrix representation just multiplication by the η matrix. So to invert the transformation and get V back, we just multiply by the inverse η^{-1}. Multiplying it by the inverse is effectively getting its index up again. We thus see that η^{-1} is the tensor that raises the indices, and we define it to be $eta^{\mu\nu}$ (with its indices up).

- So $\eta\eta^{-1} = I \xrightarrow{\text{Einstein notation}} \boxed{\eta^{\alpha\nu}\eta_{\nu\beta} = \delta^\alpha_{\ \beta}}$, where $\delta^\alpha_{\ beta}$ is just the identity matrix in einstein's notation (and it gets a special fancy name for some reason "Kronecker delta". But it's really just the identity matrix).

- So now we have that:

$$\text{To lower indices: } \boxed{V_\mu = \eta_{\mu\nu} V^\nu}$$
$$\text{To raise indices: } \boxed{V^\mu = \eta^{\mu\nu} V_\nu}$$

- Once we have decided on this, one can use the minkowski metric to even raise or lower some indices of higher rank tensors. For example: $T^{\mu\nu\ \ \alpha}_{\ \ \ \beta}\eta_{\mu\gamma}\eta^{\beta\xi} = T_\gamma^{\ \nu\xi\alpha}$

- From now on, we'll use einstein's notation in our equations of general relativity as they're simply much more convenient.

3.6 The Geodesic Equation from the einstein equivalence principle

- Given coordinates x^μ, a new coordinate system is in the most general sense a function of the old coordinates; i.e. $\xi^\mu = \xi^\mu(x^\mu)$.
- So the einstein equilvalence principle tells one that there always exists a coordinate system for a point p where it's locally flat (has the minkowski metric, and thus inertial and thus acceleration is zero). So we have that $\frac{d^2\xi^\alpha}{d^2\tau} = 0$
- Let's work this out further:

$$\frac{d^2\xi^\alpha}{d^2\tau} \overset{\text{chain rule}}{=} \frac{d}{d\tau}\left(\frac{\partial\xi^\alpha}{\partial x^\mu}\frac{\partial x^\mu}{\partial\tau}\right) \overset{!}{=} 0$$
$$\overset{\text{product rule}}{=} \frac{\partial^2\xi^\alpha}{\partial x^\mu \partial x^\nu}\frac{\partial x^\mu}{\partial\tau}\frac{\partial x^\nu}{\partial\tau} + \frac{\partial\xi^\alpha}{\partial x^\mu}\frac{\partial^2 x^\mu}{\partial\tau^2} = 0$$

Now let's multiply through out by $\frac{\partial x^\lambda}{\partial\xi^\alpha}$. Note that this is the inverse of the so called "Jacobian matrix" (matrix of derivatives) of ξ.

$$\begin{aligned}
0 &\overset{!}{=} \left(\frac{\partial x^\lambda}{\partial\xi^\alpha}\right)\frac{\partial^2\xi^\alpha}{\partial x^\mu \partial x^\nu}\frac{\partial x^\mu}{\partial\tau}\frac{\partial x^\nu}{\partial\tau} + \frac{\partial\xi^\alpha}{\partial x^\mu}\left(\frac{\partial x^\lambda}{\partial\xi^\alpha}\right)\frac{\partial^2 x^\mu}{\partial\tau^2} \\
&= \left(\frac{\partial x^\lambda}{\partial\xi^\alpha}\right)\frac{\partial^2\xi^\alpha}{\partial x^\mu \partial x^\nu}\frac{\partial x^\mu}{\partial\tau}\frac{\partial x^\nu}{\partial\tau} + \delta^\lambda_\mu\frac{\partial^2 x^\mu}{\partial\tau^2} \\
&= \left(\frac{\partial x^\lambda}{\partial\xi^\alpha}\frac{\partial^2\xi^\alpha}{\partial x^\mu \partial x^\nu}\right)\frac{\partial x^\mu}{\partial\tau}\frac{\partial x^\nu}{\partial\tau} + \frac{\partial^2 x^\lambda}{\partial\tau^2}
\end{aligned}$$

Define $\Gamma^\lambda_{\mu\nu} := \left(\frac{\partial x^\lambda}{\partial\xi^\alpha}\frac{\partial^2\xi^\alpha}{\partial x^\mu \partial x^\nu}\right)$. This is just a group of coeffecients that we call the "connection" or the "Christoffel symbols". Our final equation becomes:

$$\boxed{\frac{\partial^2 x^\lambda}{\partial\tau^2} = -\Gamma^\lambda_{\mu\nu}\frac{\partial x^\mu}{\partial\tau}\frac{\partial x^\nu}{\partial\tau}}$$

that's the so called **geodesic equation**, and it's the equation of motion of general relativity.

- On the left handside is the acceleration in the given coordinate system. Notice how it's not equal to zero. It's analogous to the coriolis effect where acceleration can emerge simply by going to a different frame of reference. We will later show that the connection Γ only depends on the metric and its derivatives; and thus that curvature induces acceleration described by this equtaion.