PS-523
Astrophysics, Gravitation and Cosmology

School of Physical Sciences,
JNU

Pankaj Sharan

Formerly Professor
Physics Department, Jamia Millia Islamia
New Delhi

January 2017

Lecture 1 : Introductory Remarks

The constant 'c' entered physics as the ratio of two systems of units. Lorentz transformations do not actually require the postulate of constancy of velocity of light for their derivation. The essential part of special relativity is the relativity of simultaneity, which leads to revision of our views on spacetime. The space time of special relativity is seen most clearly in Minkowski diagrams. The four-dimensional spacetime and the group character of Lorentz transformations requires that all physical quantities be either scalars, or 4-vectors or tensors etc.

Equivalence Principle (1907) allows simple conclusions like slowing down of clocks in gravitational field. The principle was an important step towards search for a relativistic theory of gravitation.

§ 1 : 'c' for 'constant'

The letter 'c' which we identify with the speed of light entered physics as a constant denoting the ratio of two ways in which electro-magnetic units can be defined. Two equally charged bodies placed at a unit distance repel each other by electrostatic force. The (equal) amount of charge (in the two bodies) which will produce a unit force was called one electrostatic unit (esu) of charge. This charge therefore has dimensions (the dimension of a physical quantity is denoted by enclosing its name in square brackets)

$$[\text{esu of charge}] = \sqrt{\text{force}} \times \text{distance}$$

On the other hand two parallel currents separated by a line normal to them attract each other by an inverse square law (Ampere's law) where the force is proportional to currents as well as length of current elements (say dl)

$$\text{force} \propto \frac{(\text{current})^2 \times (dl)^2}{(\text{distance})^2}.$$

Therefore, an electro-magnetic unit (emu) of current is defined as that which will produce a unit force at unit distance according to this formula. An emu of charge then is the charge that passes in unit time by one emu of current. This charge has dimension

$$[\text{emu of charge}] = \sqrt{\text{force}} \times \text{time}.$$

The ratio of the two units is a constant 'c' of the dimensions of velocity. Since we cannot use units of different dimensions for the same physical quantity, the constant 'c' hangs about in equations of electromagnetism when we choose one or the other system. Magnetic effects of currents are much, much smaller than electrostatic effects, therefore one emu of charge is a huge number when expressed in esu's.

The charge in a big condenser was measured by two different methods by Weber and Kohlrausch in 1856 to determine the constant c. A big condenser, called 'Leyden Jar' in those days, (and there is a story why it was called a Leyden jar) was charged and its charge was measured by electrostatic methods as well as by electromagnetic methods by discharging the condenser and producing a current. Their value was

$$c = 3.1 \times 10^{10} \text{ cm/s}.$$

Kirchhoff immediately noticed its near equality to the known value of speed of light and gave a theory for propagation of electric fields in perfect conductors with this velocity. Maxwell's paper (1864) "A dynamical theory of electromagnetic field" finally identified the constant 'c' as the speed of propagation of all electromagnetic fields.

Remark 1 Reference for this section Edmond Whittaker, *History of Theories of Aether and Electricity*, Vol 1, p.232, Humanities Press New York, 1973. The accidental discovery of the Leyden jar in 1745 by Professor Muschenbroeck of Leyden can be found in the book on page 45.

§ 2 : Lorentz transformations

It is generally believed that the derivation of Lorentz transformations (which were known since 1892) involves two 'basic postulates' of special relativity: namely the 'principle of relativity' which requires the equivalence of all inertial frames and the principle of 'the constancy of the velocity of light'. This later assumption is not really necessary. This has been known since the earliest days (Ignatowsky, 1910; Frank and Rothe, 1911; and rediscovered several times later by scores of people), that the following three assumptions (all of which follow from the equivalence of inertial frames), are sufficient:

1. velocity of one frame relative to another is equal and opposite to that of the latter with respect to the former
2. transformations in the same direction form a group, that is, two successive transformations in the same direction produce a similar transformation between the first and last frames
3. the length of a rod when attached to a frame when measured in another frame should be the same as when it is attached to the second frame and measured by the first.

Here is a simple proof I worked out for this lecture.

Let (x, t) and (x', t') be the two frames S and S' with S' moving with velocity v along x axis with respect to S.

Galilean transformations are inadequate because they do not respect equations of electromagnetic phenomena. The generalized choice is

$$\begin{pmatrix} x' \\ t' \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} x \\ t \end{pmatrix}$$

where A, B, C, D are functions of the relative velocity v, that is $A = A(v)$ etc. Since the origin of S' ($x' = 0$) moves with velocity v we must have $B = -vA$, therefore the form is

$$\begin{pmatrix} x' \\ t' \end{pmatrix} = \begin{pmatrix} A & -vA \\ C & D \end{pmatrix} \begin{pmatrix} x \\ t \end{pmatrix}.$$

The inverse transformation is

$$\begin{pmatrix} x \\ t \end{pmatrix} = \begin{pmatrix} A^- & vA^- \\ C^- & D^- \end{pmatrix} \begin{pmatrix} x' \\ t' \end{pmatrix}$$

where $A^- = A(-v)$ etc.

Length of a rod is the difference between the coordinates of the ends of the rod *at the same time* in the frame in which measurement is being made. So a rod at rest in S' with ends at x'_1 and $x'_2 = x'_1 + L$ will be seen for $t_1 = t_2$ in frame S to have length L/A. Conversely the same rod when placed in frame S will be seen by S' to have length L/A^-. Therefore, we must have

$$A = A^-.$$

The inverse transformation can be directly calculated by inverting the matrix as

$$\begin{pmatrix} x \\ t \end{pmatrix} = \frac{1}{A(D + vC)} \begin{pmatrix} D & vA \\ -C & A \end{pmatrix} \begin{pmatrix} x' \\ t' \end{pmatrix}.$$

Comparison shows that $A = D = A^-$. We can pull A out in the matrix and the form is

$$\begin{pmatrix} x' \\ t' \end{pmatrix} = A \begin{pmatrix} 1 & -v \\ E & 1 \end{pmatrix} \begin{pmatrix} x \\ t \end{pmatrix}.$$

where $E = C/A$ is another function of v.

Let a third frame S'' move with velocity v^* with respect to S' in the common direction. Then the frame S'' and S are related by the matrix ($A^* = A(v^*)$ etc.)

$$AA^* \begin{pmatrix} 1 & -v \\ E & 1 \end{pmatrix} \begin{pmatrix} 1 & -v^* \\ E^* & 1 \end{pmatrix} = AA^* \begin{pmatrix} 1 - vE^* & -v - v^* \\ E + E^* & 1 - v^*E \end{pmatrix}$$

In order to have the same standard form, $vE^* = v^*E$, so that

$$\frac{v}{E} = \frac{v^*}{E^*} = \text{constant independent of } v \text{ or } v^* \text{ say} = \alpha.$$

The transformation connecting the two frames S and S' therefore has the form

$$A \begin{pmatrix} 1 & -v \\ v/\alpha & 1 \end{pmatrix},$$

whose inverse is

$$\frac{1}{A} \frac{1}{(1 + v^2/\alpha)} \begin{pmatrix} 1 & v \\ -v/\alpha & 1 \end{pmatrix}$$

which should be equal to

$$A^- \begin{pmatrix} 1 & v \\ -v/\alpha & 1 \end{pmatrix} = A \begin{pmatrix} 1 & v \\ -v/\alpha & 1 \end{pmatrix}.$$

This determines

$$A = \frac{1}{\sqrt{1 + v^2/\alpha}}.$$

This is the most general form. When α is a very large number these become the Galilean transformations. Experimental determination (of electromagnetic phenomena for example) reveals that the constant (which has the dimensions of velocity square) is given by

$$\alpha = -c^2$$

where c is the electromagnetic constant or the velocity of light.

Remark 2 Reference for historical comments in this section, W. Pauli, *Theory of Relativity*, section 4, Pergamon Press 1958.

§ 3 : Relativity of Simultaneity

The Lorentz transformations reveal the structure of space-time, as was discovered by Hermann Minkowski in 1908.

The Lorentz transformations

$$x' = \gamma(x - vt),\ t' = \gamma(t - vx/c^2),\ y' = y,\ z' = z$$

where $\gamma = 1/\sqrt{1 - v^2/c^2}$ can be depicted in a graphical form symmetrically if we define $x^0 = ct, x^1 = x, x^2 = y, x^3 = z$ and similarly for t', x', y', z'. with these

$$\begin{aligned}(x')^1 &= \gamma(x^1 - \beta x^0),\ (x')^0 = \gamma(x^0 - \beta x^0),\\ (x')^2 &= x^2,\ (x')^3 = x^3\end{aligned}$$

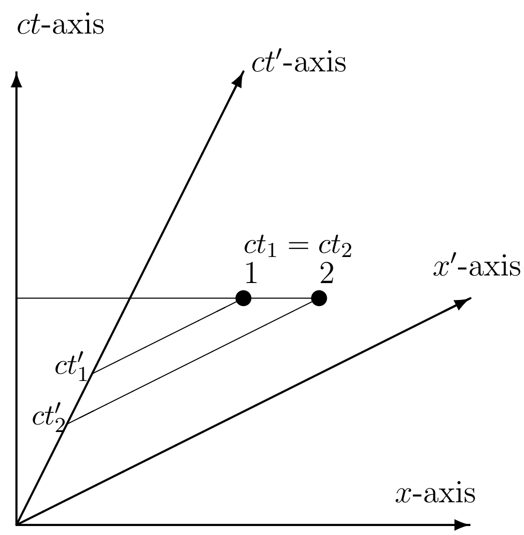
where $\beta = v/c$.

For simplicity let us consider only x^0 and x^1. In a plane we can take any two nonparallel straight lines as the x^0 and x^1 axes with the point of intersection as the origin. The x^0-axis is the straight line $x^1 = 0$ and the x^1-axis is the line $x^0 = 0$.

This allows all 'world-events' to be shown on the plane using a grid of lines parallel to the two axes.

We show the x^0 and x^1 - axes as orthogonal here, for convenience but that is not necessary at all! Any two non-parallel lines are good enough.

Similarly the $(x')^0$-axis is the straight line $(x')^1 = 0$ or $x^1 = \beta x^0$. And the $(x')^1$-axis is the line $(x')^0 = 0$ or $x^1 = \beta x^0$. If β is positive (it ia always less than one), the $(x')^0$ and $(x')^1$ axes are as shown. For a signal of light (starting from the origin) in both directions $x^0 = \pm x^1$ as well as $(x')^0 = \pm(x')^1$. These are the bisector lines of all pairs of axes. If we add x^2, x^3 coordinates as well, then the light signals will trace out a surface in four dimensions called the *light cone*.



Relative Simultaneity:

Events 1 and 2 are simultaneous for one frame $(t_1 = t_2)$ are not so for another frame $(t'_1 \neq t'_2)$

We notice the **relativity of simultaneity**. Two events in space-time which are simultaneous for (x, t) frame (that is they have the same value for the time coordinate) are not so in the other (x', t') frame which moves with respect to this frame.

Consider the trajectory of a particle moving in space time. For two neighbouring events (x_1, ct_1) and $x_2, ct_2)$ on its trajectory in the frame $(x, , ct)$ define

$$s^2 = (x_2 - x_1)^2 - c^2(t_2 - t_1)^2$$

It is a consequence of Lorentz transformations that the similar quantity has the *same* value for the (x', t') frame as well

$$s^2 = (x'_2 - x'_1)^2 - c^2(t'_2 - t'_1)^2$$

Thus this quantity is the **invariant distance** or **invariant interval** between the two events independent of the inertial frame of reference in which it is calculated.

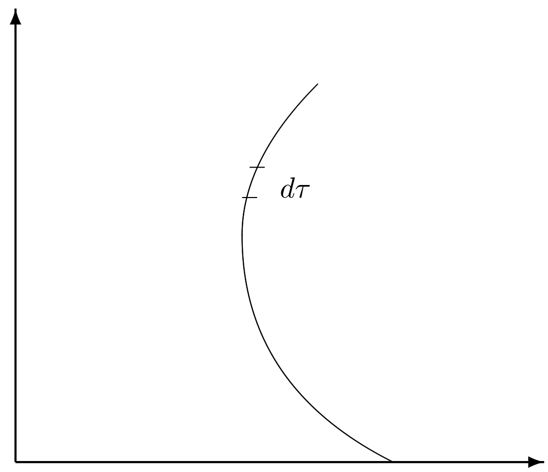
Because the speed of all objects is less than that of light s^2 is actually negative. Define $\tau = \sqrt{-s^2/c^2}$ or,

$$\tau_{21} = (t_2 - t_1)\sqrt{1 - v^2/c^2} = (t'_2 - t'_1)\sqrt{1 - v'^2/c^2}$$

where $v = (x_2 - x_1)/(t_2 - t_1)$ and $v' = (x'_2 - x'_1)/(t'_2 - t'_1)$ are the velocities of the particle as determined by these two infinitesimally close events. Because τ_{21} has the same value in all frames, it is equal to its value in the frame in which the particle is at rest $(v = 0)$. Thus τ_{21} *is the time recorded by a clock carried along with the particle.*

A particle moves with a non-uniform velocity in general. We define the proper time along the trajectory of the particle as

$$\int d\tau = \int \sqrt{1 - v^2/c^2}\, dt$$



Proper time:

$$\int d\tau = \int \sqrt{1 - v^2/c^2}\, dt$$

is the time shown by a clock traveling with the particle

Any two observers who start from the same point, go in different directions and later come again and compare their watches, will find them not agreeing in general. This is no different that the fact that two curves on a plane which start and finish at the same two points have different lengths in general.

§ 4 : Minkowski diagram

(Minkowski was a mathematician and a teacher in the same ETH in Zurich where Albert Einstein studied from 1896 to 1900. But Einstein hardly ever attended Minkowski's classes, so Minkowski was quite surprised to learn that one of his former students had discovered relativity.)

A point in Minkowski space-time is called an **event**. It is specified by the spatial coordnates $\mathbf{x} = (x^1, x^2, x^3)$ of the point and the time t. We denote by $x^\mu = (x^0 = ct, x^1, x^2, x^3)$ the coordinates of an event where the time coordinate t is multiplied by the light velocity c in order to get greater symmetry in the formulas. It also makes all the four quantities have the same physical dimension of length.

This four dimensional space, called the **Minkowski space** differs from the usual three dimensional Euclidean space in one crucial way. The length squared between two points in three dimensional Euclidean space is given by the positive definite expression

$$(\mathbf{x} - \mathbf{y}).(\mathbf{x} - \mathbf{y}) = (x^1 - y^1)^2 + (x^2 - y^2)^2 + (x^3 - y^3)^2$$

which is always positive. In the Minkowski space the square of the interval is given by

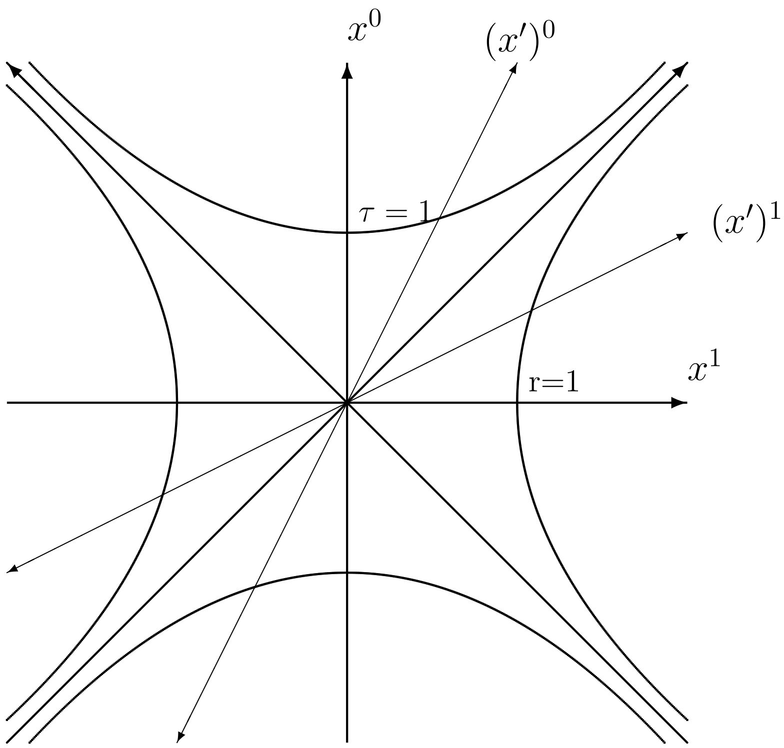
$$\begin{aligned}
(x - y).(x - y) &= -(x^0 - y^0)^2 + (x^1 - y^1)^2 + (x^2 - y^2)^2 + (x^3 - y^3)^2 \\
&= -(x^0 - y^0)^2 + (\mathbf{x} - \mathbf{y}).(\mathbf{x} - \mathbf{y}) \\
&= \eta_{\mu\nu}(x^\mu - y^\mu)(x^\nu - y^\nu)
\end{aligned}$$

where we use summation over indices $\mu, \nu = 0, .., 3$. This expression can be positive, negative or zero. The numbers $\eta_{\mu\nu}$ are elements of a matrix η which are equal to $-1, +1, +1, +1$ on the diagonal and zero elsewhere.

The important point is that this **interval** between two space time events x and y is *invariant*, that is a number which has the same value in all coordinate frames.

Each inertial frame of reference is represented by a choice of axes in this space. Lorentz transformations between inertial frames are 4×4 matrices Λ which connect the coordinates x^0, x^1, x^2, x^3 and x'^0, x'^1, x'^2, x'^3 corresponding to the *same physical event* (written as column matrices) : $x' = \Lambda x$

which must satisfy the condition that the interval has the same invariant value $s^2 = (x-y).(x-y) = (x'-y').(x'-y')$ for a pair of events whose coordinates in one frame are x nd y and in the other x' and y'.



Minkowski Diagram

The $(x')^0$-axis corresponds to $x^1 - \beta x^0 = 0$, and the $(x')^1$-axis corresponds to $x^0 - \beta x^1 = 0$. The hyperbolas correspond to points at constant space-like and constant time-like distances from the origin. The 45° lines denote the light cone.

Between any two infinitesimally close space-time events with coordinates x^μ and $x^\mu + dx^\mu$ the infinitesimal interval can be written

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

When the interval is positive the events are called **space-like separated**, when it is negative they are **time-like separated events**, and when it is zero they are **light-like separated**. If a point particle is present at x and also at $x + dx$ then (assuming $dt > 0$) its velocity $\mathbf{v} = d\mathbf{x}/dt$ where $\mathbf{x} = (x^1, x^2, x^3)$, is related to the invariant interval as

$$(ds)^2 = -c^2(dt)^2 + (d\mathbf{x})^2 = -c^2(dt)^2 + \mathbf{v}^2(dt)^2 = -(c^2 - \mathbf{v}^2)dt^2$$

As the particle cannot have velocity greater than that of light the interval is time-like. We call

$$d\tau = \frac{1}{c}\sqrt{-\eta_{\mu\nu} dx^\mu dx^\nu} = dt\left(1 - \frac{\mathbf{v}^2}{c^2}\right)^{\frac{1}{2}}$$

the **proper time** between the particle's two successive space-time positions. The invariant number $d\tau$ is the time as seen by an observer in the rest frame of the particle, because for a clock traveling with the particle, the velocity $\mathbf{v}$ is zero.

For an arbitrary motion of the particle, the sum of infinitesimal proper times along its four dimensional trajectory will be the time shown by a clock we can imagine traveling along with the particle. Two clocks starting from the same initial event and showing the same time at that event can follow separate trajectories and may meet again at a common space-time point later on. But on comparison will show different times in general. The integrated proper-times along the two trajectories need not be equal. This is no more surprising than the fact that arc-lengths of two arbitrary curves between two fixed points in Euclidean plane are different in general.

In fact the proper time plays the same role as the length of a curve in Euclidean space.

The motion of a mass point can be described by a curve $x(\tau)$ in space-time, where τ is the proper time labelling points on the trajectory curve. For a free particle without any force acting on it the trajectory is a time-like straight line in space-time :

$$x^\mu(\tau) = x^\mu(0) + \tau U^\mu, \qquad \frac{dx^\mu}{d\tau} = U^\mu, \qquad \frac{d^2x^\mu}{d\tau^2} = 0$$

A time-like straight line in space-time between two fixed points is a **geodesic**, that is, a path for which the integral of ds along the path is extremal. A variational principle can be written with action A

$$\delta A = -mc^2\delta \int d\tau = -mc\delta \int \sqrt{-\eta_{\mu\nu}dx^\mu dx^\nu} = 0$$

This looks a little more familiar if we rewrite it in non-relativistic limit of small velocities $|\mathbf{v}| << c$

$$A = -mc^2 \int d\tau = -mc^2 \int dt \left(1 - \frac{\mathbf{v}^2}{c^2}\right)^{\frac{1}{2}} \approx -mc^2(t_2 - t_1) + \frac{1}{2}\int m\mathbf{v}^2 dt$$

where the first term depends only on the endpoints and the second term shows the kinetic energy as the Lagrangian of a the free non-relativistic particle.

In four dimensions, *dynamics* becomes *geometry*. The force free motion becomes straight lines or extremal paths given by $\delta \int d\tau = 0$.

Light signals also move along straight lines but the proper time along the line is zero.

§ 5 : Four dimensional quantities

The main lesson of Minkowski's contribution to special theory of relativity is that that physical quantities have a four dimensional character. Quantities which were described by three-vectors (that is vectors with three components) before the theory of relativity are seen to have a fourth partner corresponding to the time-axis. Such quantities were called **four-vectors**.

For example the "four-velocity" of the particle U is defined as the derivative with respect to the proper time : it is the tangent vector to the trajectory.

$$U^\mu = \frac{dx^\mu}{d\tau} = (c\gamma(|\mathbf{v}|), v^i\gamma(|\mathbf{v}|)) \tag{1}$$

where $v^i \equiv dx^i/dt$ the usual three-velocity of the particle. Note that the three-velocity is not just the "space" part of the velocity four vector. Note also, that all four components of the velocity four-vector are not independent

$$< U, U >= \eta_{\mu\nu}\frac{dx^\mu}{d\tau}\frac{dx^\nu}{d\tau} = -c^2 \tag{2}$$

We can say that the *four-velocity of a material particle is a time-like vector of constant magnitude.*

At any (proper) time τ there exists a coordinate system in which the particle is momentarily at rest, that is, its velocity four-vector has components $(c, 0, 0, 0)$.

§ 6 : Equivalence Principle

Imagine a box the size of a small room with an observer inside it resting on Earth's surface. The observer cannot see out of the box. Sitting inside she can infer that there is a gravitational field by observing that bodies in the box fall with the same acceleration **g**.

Now take this box (alongwith the observer) to a place far away from gravitating bodies like the Sun or Earth and (with respect to any inertial frame) give a constant acceleration to the box.

To the observer inside everything would again be seen to fall with acceleration $\mathbf{g}$ and there would be no way of knowing that she is not, in fact, in a constant gravitational field.

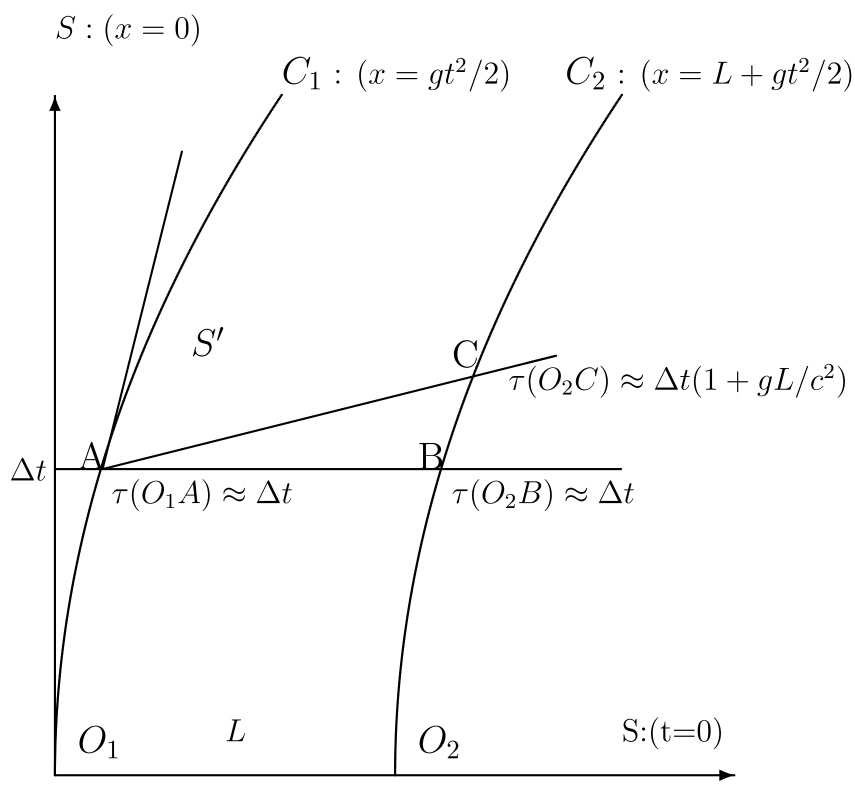
This equivalence of constant gravitational field with the physical effects in an accelerated frame was called by Einstein the **Equivalence Principle**. He made an even stronger assumption that one would not be able to ditinguish between the two situations *by any physical experiments* (and not merely the mechanical experiments used for measuring the accelerations of bodies). Sometimes this is called the Strong Principle of equivalence.

An immediate consequence of the principle is that an observer falling freely in a constant gravitational field would be in an inertial frame. The freely falling observer experiences *no gravitational field !* Every object around her will fall with the same acceleration under gravity and there would be no motion ascribable to gravity.

Using special relativity it was possible to find physical effects in an instantaneous frame "comoving" with an accelerated frame, and by using the Equivalence principle one could therefore guess what would happen in a gravitational field. For example, Einstein proved in 1907 that clocks near a massive body (where gravitational potential is lower) run slowly compared to clocks farther away.

We follow Einstein's argument of 1907 to show that clocks run slowly where gravitational potential is low compared to clocks at higher value of potential.

Let there be an inertial frame S with respect to which another frame S_1 is at rest at $t = 0$ with coinciding axes and the origin. Let S_1 start accelerating in the x-direction at $t = 0$ with acceleration g.



Einstein's argument of 1907

Clocks run slowly at lower gravitational potential.

Let there be two clocks C_1 and C_2 in the accelerating frame S_1. They are at events O_1 and O_2 at $t = 0$ as seen by S. As shown the trajectories of the clocks C_1 and C_2 according to frame S are given by $x_1(t)$ and $x_2(t)$ where

$$x_1(t) = \frac{1}{2}gt^2, \qquad x_2(t) = L + \frac{1}{2}gt^2$$

These equations describe the accelerated frame at low enough velocities, that is, at small values of t.

It is clear from the diagram that the clock C_2 (at event C) will show more time that the clock C_1 at A. Actually,

$$\tau(O_2C) \approx \tau(O_1A)(1 + gL/c^2)$$

Therefore, in the accelerated frame, which feels a constant gravitational field in $-x$ direction , the clock at 'height' L shows a reading $\Delta t(1 + \Phi/c^2)$ when at the same instant (according to this frame) the clock at the origin shows a reading Δt. Here $\Phi = gL$ is the gravitational potential difference between the locations of the clocks.

Similarly one can prove that light rays moving in a gravitational field bend from rectilinear path in the direction of the gravitational field.

These are preliminary results of an incomplete theory, but were crucial for progress.

Gravitational fields can be appproximated to be uniform only in very small regions of spacetime. The equivalence principle therefore holds only in such regions. Fortunately, laws of physics are expressed in terms of local differential equations for quantities like fields. Therefore the priciple can be used to generalise laws from their formulations in a 'freely falling local inertial frame' to arbitrary gravitational fields.

The equivalence principle proved to be the guiding principle for finding the general equations. One was required to put (*in a suitable form*) the equations in the inertial freely falling frame where we knew the special theory of relativity holds, and then declare them to be valid for all frames of reference.

§ 7 : Need for the Riemannian geometry

There are three physical questions to be answered for a relativistic theory of gravity : (1) which quantities describe the gravitation field? (2) what are the equations that relate the gravitational field to matter-energy distribution? and (3) what is the equation of motion of a particle in the gravitational field.

In the Newtonian theory the gravitational potential $\Phi(x)$ describes gravity. At any instant the matter distribution is given by the mass density ρ. This density determines the potential at the same instant of time by the Poisson equation

$$\nabla^2\Phi = 4\pi\rho$$

In this gravitational field a mass point moves along a trajectory determined by

$$\frac{d^2\mathbf{x}}{dt^2} = -\nabla\Phi$$

It looks natural to generalise Φ into a scalar field to describe gravity in the relativistic case. But such a theory proposed by Einstein and others between 1908-1912 does not work.

The first clues came from the equivalence principle.

According to the Equivalence Principle, there must exist in a small region of space-time a coordinate system X^0, X^1, X^2, X^3 in which the equation of the particle is a straight line i.e. force free.

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

This equation can be derived from the variational principle $\delta \int d\tau = 0$. In a general coordinate system, the same equation holds because $d\tau$ is invariant :

$$\begin{aligned} c^2(d\tau)^2 &= -(ds)^2 \\ &= (dX^0)^2 - (dX^1)^2 - (dX^2)^2 - (dX^3)^2 \\ &= -g_{\mu\nu} dx^\mu dx^\nu \end{aligned}$$

where $g_{\mu\nu}$ are determined by the relation between the coordinates X^μ and x^μ

$$g_{\mu\nu} = \eta_{\alpha\beta} \frac{\partial X^\alpha}{\partial x^\mu} \frac{\partial X^\beta}{\partial x^\nu}$$

In a general frame the particle will be seen moving under the influence of gravitational forces, along paths determined by $\delta \int d\tau = 0$ where proper time $d\tau$ is given by the expression in terms of $g_{\mu\nu}$. Therefore $g_{\mu\nu}$ must be intimately related to the gravitational field. The first question (about which quantities represent gravitational field) was thus resolved.

This expression for $(d\tau)^2$ is the familiar 'line element' in the Gauss-Riemann geometry. It became clear that the Riemannian geometry is the proper tool for gravitational theory.

Working along these lines, Einstein collaborated with his mathematician friend Marcel Grossmann. The equation $\delta \int d\tau = 0$ was seen to lead to the equation for the 'straightest curve' or the geodesic

$$\frac{d^2 x^\mu}{d\tau^2} = -\Gamma^\mu_{\nu\sigma} \frac{dx^\nu}{d\tau} \frac{dx^\sigma}{d\tau}$$

where $\Gamma^\mu_{\nu\sigma}$ are expressions in terms of derivatives of $g_{\mu\nu}$. These quantities will be defined later.

In the non-relativistic limit, this equation reduces to the Newtonian equation if we take, for $(x^0 = ct)$

$$g_{00} = -\left(1 + \frac{2\Phi}{c^2}\right)$$

From the form of the equations for the particle, Γ appear to be related to gravitational force and the $g_{\mu\nu}$ play the role of the gravitational potential. This takes care of the third question above.

The search for the *field equations* which determine $g_{\mu\nu}$, from a knowledge of the energy-matter distribution, (that is, the second question in our list) took the greatest effort. After many false steps Einstein finally arrived at the correct theory in November 1915.

§ 8 : General Theory of Relativity

As we have seen, it is the geometry of space *and* time which departs from the Eulidean. The Minkowski space of special relativity has a non-Euclidean metric or line element.

Einstein's final version of the theory regards the spacetime as a four dimensional Riemannian space or manifold.

Let the neighbouring points have coordinates $x = (x^0, x^1, x^2, x^3)$ and $x + dx = (x^0 + dx^), x^1 + dx^1, x^2 + dx^2, x^3 + dx^3)$ where x^0 is a 'time' coordinate and $x^i, i = 1, \dots, 3$ refer to space. The infinitesimal distance squared is given by an expression

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

where indices $\mu, \nu = 0, \dots, 3$. Different coordinates systems can be used to specify the same space-time points, but the quantity ds^2 remains the same. This invariance of ds^2 determines how the metric tensor components change from one coordinate system to other.

The special theory of relativity as interpreted by H. Minkowski in 1908 uses a spacetime in which the metric tensor

$$g_{\mu\nu} = \eta_{\mu\nu}$$

has non-zero components

$$\eta_{00} = -1, \qquad \eta_{11} = \eta_{22} = \eta_{33} = 1$$

Lorentz transformations are precisely those coordinate changes in which these constant values of $\eta_{\mu\nu}$ remain the same. A spacetime continuum in which it is possible to choose a coordinate system which makes $g_{\mu\nu}$ constant everywhere like here is called **flat** or **Minkowskian**.

In the presence of gravitating matter, $g_{\mu\nu}$ get modified, and it is not possible to choose coordinate systems in which they can take constant values. The space becomes **curved**, and the measure of curvature is the Riemann-Christoffel **curvature tensor**. Its components $R_{\mu\nu\sigma\tau}$ depend on $g_{\mu\nu}$ and their derivatives up to second order.

The curvature tensor is the true measure of the 'real' gravitational field that cannot be transformed away by a choice of coordinates. Even so, in a very small neighbourhood of spacetime point one can choose a coordinate system such that the derivatives of $g_{\mu\nu}$ vanish at the point. Therefore those effects which depend on the first derivatives of the $g_{\mu\nu}$ (like the acceleration of a particle falling in gravity) are indistinguishable from that of a particle moving in gravity-free region. This is the physical content of the Equivalence Principle.

The $g_{\mu\nu}$'s are determined by energy and matter distribution by the **Einstein Equation**

$$G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = \frac{8\pi G}{c^4} T_{\mu\nu}$$

where the **Einstein tensor** $G_{\mu\nu}$ is a simple combination of the so called **Ricci tensor** given by $R_{\mu\nu} = R^{\sigma}{}_{\mu\sigma\nu}$ and the **scalar curvature** $R = g^{\mu\nu} R_{\mu\nu}$.

c and G are the velocity of light and the Newton's gravitational constant G. The constant multiplying the the right hand side is chosen so that in the non-relativistic limit the theory gives the Newton's law of gravitation.

The right hand side, contains $T_{\mu\nu}$ the **stress-energy tensor** of matter which contains information about momentum and energy densities and pressure of matter and radiation.

Material particles move in the gravitational field determined by $g_{\mu\nu}$ along **geodesics**, the "straightest possible" curves. The equation for such curves $x^\mu(\tau)$ is

$$\frac{d^2x^\mu}{d\tau^2} + \Gamma^\mu_{\nu\sigma}\frac{dx^\nu}{d\tau}\frac{dx^\sigma}{d\tau} = 0$$

where $\Gamma^\mu_{\nu\sigma}$ called Christoffel symbols are functions of $g_{\mu\nu}$ and their derivatives.

Light rays also move along geodesics except that along the path the interval $d\tau$ between any two neighbouring points is zero.

It must be remembered that curves corresponding to motion of a body in a gravitational field are the straightest possible in *space-time* and not in the three dimensional space. Thus a particle thrown vertically upwards in Earth's gravity may seem to have a highly curved path at its turning point in the three dimensions of space, but it is as straight a path as can be in the four dimensional space-time. This spectacular 'straightening' happens when we go from the three dimensional projection of the trajectory to the actual curve in four dimensions because the speed of light has a very large value compared to ordinary velocities.

General theory of relativity is a theory of gravitation in the narrow sense that if there was no gravity then special theory of relativity would suffice.

In the broad sense it is a fundamental theory which includes all physical phenomena because gravity acts on everything. As gravity is a weak force at ordinary distances and mass densities, the corrections to Newton's law of gravitation are small. The typical corrections to Newtonian theory are of the order of the dimensionless number GM/rc^2 where M is the mass of the gravitating body, r the typical distance scale of the problem.

For example, a clock at a distance r from a spherical body of mass M runs in a ratio $\sqrt{1 - 2GM/rc^2}$ slower than a clock very far away from the body. This leads to frequencies of spectral lines emitted by atoms, (atoms act like a clock), at a distance r away from the massive body to be reduced by a fraction GM/rc^2 when seen by a far away observer. This is called gravitaional red-shift (because in the visible spectrum lower frequencies occur near the red). For the atom sitting on the surface of the Sun this comes to be of the order of 10^{-6}.

The angle of bending of light from a distant star grazing past the surface of the Sun is $4GM/Rc^2$ where R is the radius of the Sun.

Similarly, the angle by which the elliptical orbit of a planet fails to close in one revolution comes out in Einstein's theory to be $6\pi GM/Lc^2$ where M is the mass of the Sun, and L, the latus rectum of the orbit. For the innermost planet Mercury (which has the smallest value of L) this angle comes out to be equal to $43''$ of arc in a hundred years (415 revolutions of the planet). This "precession of the perihelion" was the observed, but unexplained, leftover discrepancy in the orbit of Mercury after all known causes for this phenomenon had been taken into account. This was a great triumph for the general theory of relativity.

Gravitational disturbances, which mean perturbations in the values of $g_{\mu\nu}$ caused by changing matter distribution travel out as gravitational waves, taking away energy. However the direct evidence of gravitational waves is yet to come. An ambitious world-wide program for the detection of gravitational waves is currently underway and the first results are expected soon.

One of the most dramatic application of general theory of relativity is in the **gravitational collapse** and formation of **black holes**. One finds that an idealised, extremely heavy mass concentrated at a point or very small region so that this region is inside the its **Schwarzschild radius** $R_S = 2GM/c^2$ will give rise to a gravitational field so strong that even light cannot escape from this region. Such a region is called a black hole. It is a consequence of general theory of relativity that if a sufficiently large mass distribution begins to collapse under its own weight there may be no known forces of nature to stop it. In that case the collapse goes on unchecked and the mass densities may reach such high values whose physics we do not understand. The space-time is so highly curved around the cores of such objects that light gets bent and 'sucked back in', unable to escape. There are believed to be black holes of all sizes existing in Nature, ranging in total mass from a few solar masses to blackholes of a billion solar masses residing in the centers of most galaxies.

An application of the general theory of relativity is to the behavior of the universe itself. Assuming the universe at the largest distance scales (of a few billion light years) to be uniform and isotropic it is possible to infer from Einstein's equations that the universe must have started from a hot fiery ball which has been expanding ever since. The field of **cosmology** has received a great boost in recent years due to fantastic progress in observational astrophysics.

All the same the general theory of relativity is a *classical* theory. Physicists believe that all phenomena, at the appropriate level, must be described by a quantum theory. All attempts to quantize the classical gravitational field described by Einstein's theory have been unsuccessful so far. Physicists hope to learn something deep and fundamental once they are ultimately successful in doing that.

Lecture 2 : Vector and Tensor Algebra

A vector in a vector space V can be represented by a set of numbers called its components once a basis is chosen. When a new basis is chosen, the components change by a matrix which is the inverse transpose of the matrix with which the basis was changed. Thus the components transform **contravariantly**. *The set of all linear functionals on a vector space V itself forms a vector space V^*, called* **dual** *to V, of the same dimension under obvious definitions of additional and scalar multiplication. Moreover, a basis can be defined in V^* given a basis in V in a standard way. When a basis is changed in V, the dual basis in V^* changes by the inverse transpose matrix so that components of vectors in V^* change by inverse transpose of inverse transpose of the matrix which is the original matrix. These vector components transform* **covariantly**. *Bilinear functionals on a vector space define vector spaces too, related to their tensor product. The exterior or 'wedge product' is an antisymmetric combination of tensor products.*

§ 9 : Vector Spaces

A **real vector space** V is a set whose members, called **vectors**, have two operations defined on them. Two vectors can be added to give a vector called their sum, and, a vector can be multiplied by a real number to give a vector. A vector space has a special vector $\mathbf{0}$, called the **zero vector**, which has the property that when added to any vector of the space the sum is this latter vector.

Two (non-zero) vectors are called **linearly independent** if they are not proportional to each other, that is, one can not be wrtitten as a number times the other. More generally, a set of r non-zero vectors $\mathbf{v}_1, \ldots, \mathbf{v}_r$ is called linearly independent if none of them can be written as a **linear combination** of any, some or all the others. A linear combination of vectors is just the sum of these vectors after they have been multiplied by numbers.

If vectors $\mathbf{v}_1, \ldots, \mathbf{v}_r$ are linearly independent then $a_1\mathbf{v}_1 + \cdots + a_r\mathbf{v}_r = \mathbf{0}$ implies that $a_1 = \cdots = a_r = 0$ because if any of the a's fail to be zero, let us say, $a_k \neq 0$, then, by dividing by a_k in the equation we can express $\mathbf{v}_k$ in terms of the others.

Let there be a set of linearly independent vectors. Let us adjoin to this set a new non-zero vector, then this set can be linearly independent, or it may fail to be linearly independent. We can try to make a set which is the largest possible linearly independent set by adjoining only those vectors which make the larger set linearly independent.

We will deal only with those spaces, called **finite dimensional** vector spaces, in which this process of finding the largest set of linearly independent set comes to an end and we have a finite set $\mathbf{v}_1, \ldots, \mathbf{v}_n$ of linearly independent vectors. Any other vector of the space can then be written as a linear combination of these vectors. Whichever way we choose the set of linearly independent vectors, it always contains the same number n of vectors.

A set of vectors like these is called a **basis**. The number n of independent vectors is characteristic of the space and is called the **dimension** of the space.

§ 10 : Dual vector space

A **functional** α defined on a vector space V is a mapping which assigns to each vector $\mathbf{v} \in V$ a real number $\alpha(\mathbf{v})$. If this mapping satisfies the linearity property, that is for every $\mathbf{v}, \mathbf{u}$ in V and any real number a if

$$\alpha(\mathbf{v} + \mathbf{u}) = \alpha(\mathbf{v}) + \alpha(\mathbf{u}) \tag{3}$$

$$\alpha(a\mathbf{v}) = a\alpha(\mathbf{v}) \tag{4}$$

then we call α a **linear functional**.

Now consider the set V^* of all linear functionals on the space V. We can define addition and multiplication on linear functionals by

$$(\alpha + \beta)(\mathbf{v}) = \alpha(\mathbf{v}) + \beta(\mathbf{v})$$

$$(a\alpha)(\mathbf{v}) = a\alpha(\mathbf{v})$$

With these definitions V^* becomes a vector space whose zero vector is the linear functional which assigns number zero to each vector of V. The space V^* is called the vector space **dual** to V.

§ 11 : Change of basis

Let V be a vector space of dimension n and let $E = \{\mathbf{e}_i\}_{i=1}^n$ be a basis in V. We write a vector $\mathbf{v} \in V$ in terms of the basis as $\mathbf{v} = x^i e_i$ where there is a sum over i on the right hand side from 1 to n. We shall use this tacit assumption of a sum over a repeated index without showing the sign of summation. This is called the **Einstein summation convention**. Any exception to the convention will be explicitly pointed out.

The numbers $x^i, i = 1, \dots, n$ are called **components** of vector $\mathbf{v}$ with respect to basis E. We are using another convention by writing the components with a superscript.

Let $F = \{\mathbf{f}_i\}_{i=1}^n$ be another basis of V whose vectors can be expanded in terms of the old basis vectors as

$$\mathbf{f}_i = T_i^{\ j}\mathbf{e}_j \tag{5}$$

The numbers $T_i^{\ j}$ can be treated as elements of a matrix, with row index i and column index j. It is an invertible matrix with the inverse matrix giving coefficients involved in expanding members of basis E in terms of those of F.

Let the components of a vector $\mathbf{v}$ with respect to basis F be y^i then

$$\mathbf{v} = y^i\mathbf{f}_i = y^iT_i^{\ j}\mathbf{e}_j = x^j\mathbf{e}_j$$

Therefore,

$$y^k = x^j{(T^{-1})_j}^k = {(T^{-1})^T}^k{}_j x^j \tag{6}$$

We state the above result as follows : when the basis changes from E to F by matrix T then the components of a vector change by the matrix $(T^{-1})^T$.

§ 12 : Dual bases

Because of linearity a functional $\alpha \in V^*$ is defined completely if we give its values on a basis of V. As

$$\alpha(\mathbf{v}) = \sum x^i\alpha(\mathbf{e}_i)$$

α is defined if we know the numbers $a_i = \alpha(\mathbf{e}_i)$. This gives us an idea how to defiine a basis in V^*. Let $\alpha^i \in V^*, i = 1, \ldots, n$ be defined as

$$\alpha^i(\mathbf{e}_j) = \delta^i_j \tag{7}$$

where the Kronecker delta δ^i_j has the value 1 if $i = j$ and 0 if $i \neq j$. Any vector $\alpha \in V^*$ can be written $\alpha = a_i \alpha^i$ where $a_i = \alpha(e_i)$ as can be checked by operating α on an arbitrary vector $\mathbf{v} = v^i \mathbf{e}_i$:

$$a_i \alpha^i (v^j \mathbf{e}_j) = a_i v^j \delta^i_j = a_i v^i = v^i \alpha(\mathbf{e}_i) = \alpha(\mathbf{v})$$

This shows that $\{\alpha^i\}_{i=1}^n$ form a basis in V^*. The basis $A = \{\alpha^i\}_{i=1}^n$ is called the **basis dual** to the basis E. For every basis of V there is a dual basis in V^*.

§ 13 : Change of the dual basis

Let $B = \{\beta^i\}_{i=1}^n$ be dual to the basis F discussed above in section 5.1.3

$$\beta^i(\mathbf{f}_j) = \delta^i_j, \qquad (\mathbf{f}_i = T_i{}^j \mathbf{e}_j)$$

then B is related to A by by $(T^{-1})^T$

$$\beta^i = {(T^{-1})^T}^i{}_k \alpha^k$$

because

$$\beta^i(\mathbf{f}_j) = {(T^{-1})^T}^i{}_k \alpha^k (T_j{}^l \mathbf{e}_l) = {(T^{-1})^T}^i{}_k T_j{}^k = \delta^i_j$$

This means that if α is any general vector in V^* with components a_i in basis A : $\alpha = a_i \alpha^i$ then the components b_i with respect to basis B ($\alpha = b_i \beta^i$) will be related to a_i by the inverse transpose of $(T^{-1})^T$ that is by T itself.

§ 14 : Contra- and co-variant vectors

If we start with the space V and change basis in V by a matrix T then components of a vector in V transform **contravariantly**, that is, by $(T^{-1})^T$ whereas the components of a vector in dual space V^* (due to corresponding changes in the dual bases) transform **covariantly**, that is, by T itself.

We emphasize the vectors themselves do not change, *it is only their components that change when bases do.*

Notice the use of *superscripts* for components of vectors in V and *subscripts* for components of those in V^*. This is the standard convention of classical tensor analysis adopted by physicists.

As can be verified immediately the dual $(V^*)^*$ of V^* is V itself with the linear functional $\mathbf{v} \in V^{**} = V$ on V^* acting as

$$\mathbf{v}(\alpha) = \alpha(\mathbf{v})$$

This makes the designation of contra- and covariant quantities a matter of convention. We must decide which our starting space V is. Then vectors of V will have components transforming contravariantly, and those of V^* will have components transforming covariantly.

In our application of vectors and tensors to differential geometry fortunately there is a vector space singled out uniquely. That is the **tangent space** at any point of a differentiable manifold.

§ 15 : Tensor Product

Just as the set of linear functionals on a vector space V form a vector space V^*, the set of all *bilinear* functionals which map a pair of vectors of V into real number form an n^2 dimensional space $V^* \otimes V^*$.

Let V be a vector space of dimension n.

Consider the Cartesian product set $V \times V$. This is the set whose members are ordered pairs like $(\mathbf{v}, \mathbf{u})$ of vectors $\mathbf{v}, \mathbf{u}$ of V. This cartesian product is *just a set* and not a vector space.

A **bilinear** functional t on $V \times V$ is a mapping which assigns to each pair $(\mathbf{v}, \mathbf{w}) \in V \times V$ a real number $t(\mathbf{v}, \mathbf{w})$ with the following properties

$$t(\mathbf{u} + \mathbf{v}, \mathbf{w}) = t(\mathbf{u}, \mathbf{w}) + t(\mathbf{v}, \mathbf{w}), \qquad t(a\mathbf{v}, \mathbf{w}) = at(\mathbf{v}, \mathbf{w})$$

$$t(\mathbf{u}, \mathbf{v} + \mathbf{w}) = t(\mathbf{u}, \mathbf{v}) + t(\mathbf{u}, \mathbf{w}), \qquad t(\mathbf{v}, a\mathbf{w}) = at(\mathbf{v}, \mathbf{w})$$

The set of all bilinear mappings on $V \times V$ forms a vector space W if we define the sum of two bilinear mappings and multiplication by a number as

$$(t + s)(\mathbf{v}, \mathbf{w}) = t(\mathbf{v}, \mathbf{w}) + s(\mathbf{v}, \mathbf{w}), \qquad (at)(\mathbf{v}, \mathbf{w}) = at(\mathbf{v}, \mathbf{w})$$

We shall presently identify this space W as the n^2 dimensional space called the tensor product of the vector space V^* with itself.

§ 16 : Tensor product $T^0_2 = V^* \otimes V^*$

Let α and β be two linear functionals on V, that is, members of V^*. We can form a bilinear functional out of these as follows.

Let $\alpha \otimes \beta$ called the **tensor product** of α and β be given by:

$$(\alpha \otimes \beta)(\mathbf{v}, \mathbf{w}) = \alpha(\mathbf{v})\beta(\mathbf{w}) \tag{8}$$

It is trivial to check that this indeed is a bilinear functional. This definition also gives us properties of this tensor product.

$$\alpha \otimes (\beta + \gamma) = \alpha \otimes \beta + \alpha \otimes \gamma, \qquad \alpha \otimes (a\beta) = a(\alpha \otimes \beta) \tag{9}$$

$$(\alpha+\beta)\otimes\gamma = \alpha\otimes\gamma + \beta\otimes\gamma \qquad (a\alpha)\otimes\beta = a(\alpha\otimes\beta) \quad (10)$$

The vector $\alpha\otimes\beta$ belongs to the vector space W of all bilinear maps on $V\times V$. A vector of this type is called **decomposable** or **factorizable**. An arbitray bilinear functional on $V\times V$ is not decomposable but can always be written as a linear combination of such vectors.

A bilinear map t is completely determined by its values $t_{ij} = t(\mathbf{e}_i, \mathbf{e}_j)$ on the members of a basis E because of the linear property. It can be actually written as

$$t = t_{ij}\alpha^i\otimes\alpha^j$$

using the dual basis because t and $t_{ij}\alpha^i\otimes\alpha^j$ give the same result when acting on an arbitrary pair $(\mathbf{v}, \mathbf{u})$:

$$\begin{aligned}(t_{ij}\alpha^i\otimes\alpha^j)(\mathbf{v},\mathbf{u}) &= t(\mathbf{e}_i,\mathbf{e}_j)\alpha^i(\mathbf{v})\alpha^j(\mathbf{u}) \\ &= t(\mathbf{v},\mathbf{u})\end{aligned}$$

The last line follows because for any vector the identity $\mathbf{v} = \alpha^i(\mathbf{v})\mathbf{e}_i$ holds.

We also note from this result that the space W is n^2 dimensional and that $\{\alpha^i\otimes\alpha^j\}_{i,j=1}^n$ form a basis in it. We denote the space W by $V^*\otimes V^*$ or by T_2^0 and call it the tensor product of the space V^* with itself. Members of $T_2^0 = V^*\otimes V^*$ are called **covariant tensors of rank** 2 or (0,2) tensors. The components t_{ij} of tensor t in the basis $\{\alpha^i\otimes\alpha^j\}_{i,j=1}^n$ change to t'_{ij} in basis $\{\beta^i\otimes\beta^j\}_{i,j=1}^n$ and they are related as

$$t'_{ij} = t(\mathbf{f}_i,\mathbf{f}_j) = (T_i{}^k\mathbf{e}_k, T_j{}^l\mathbf{e}_l) = T_i{}^k T_j{}^l t_{kl} \qquad (11)$$

§ 17 : Tensor product $T_0^2 = V\otimes V$

We have already noted that just as V^* is dual to V, V is dual to V^*. Thus starting with bilinear functionals on the cartesian product $V^* \times V^*$ we can define the tensor product $\mathbf{v} \otimes \mathbf{w}$ of vectors in V exactly in the same manner as in the last section. The resulting space $V \otimes V$ is called the tensor product of spaces V. Its vectors are called **contravariant tensors of second rank** or (2,0) tensors.

A basis in the space $V \otimes V$ is given by $\{\mathbf{e}_i \otimes \mathbf{e}_j\}_{i,j=1}^n$ A tensor $t \in V \otimes V$ is completely determined by its values on (α^i, α^j) that is by numbers $t^{ij} = t(\alpha^i, \alpha^j)$. It is obvious that we can write $t = t^{ij}\mathbf{e}_i \otimes \mathbf{e}_j$. Under a change of basis from E to F by a matrix T, the components of a contravariant vector transform as

$$t'^{ij} = t(\beta^i, \beta^j) = t((T^{-1T})^i{}_k \alpha^k, (T^{-1T})^j{}_l \alpha^l) = (T^{-1T})^i{}_k (T^{-1T})^j{}_l t^{kl} \quad (12)$$

§ 18 : Multilinear functional and T_r^0

The formalism of the last sections can be generalized to multilinear functionals. A multilinear functional t on $V \times \cdots \times V$ (r-factors) is a map which assigns real number $t(\mathbf{u}, \ldots, \mathbf{w})$ to ordered set of r vectors of V, $(\mathbf{u}, \ldots, \mathbf{w})$ in such a way that

$$t(\mathbf{u} + \mathbf{v}, \ldots, \mathbf{w}) = t(\mathbf{u}, \ldots, \mathbf{w}) + t(\mathbf{v}, \ldots, \mathbf{w})$$

$$t(a\mathbf{v}, \ldots, \mathbf{w}) = at(\mathbf{v}, \ldots, \mathbf{w})$$

with similar equations for each of the arguments.

Exactly as in the bilinear case one can define $\alpha \otimes \cdots \otimes \beta$ (r-factors) as the multilinear functional

$$(\alpha \otimes \cdots \otimes \beta)(\mathbf{u}, \ldots, \mathbf{w}) = \alpha(\mathbf{v}) \cdots \beta(\mathbf{w}) \quad (13)$$

The vector space of all such multilinear functionals is called the space $T_r^0 = V^* \otimes \cdots \otimes V^*$ of **covariant tensors of rank** r or (0,r) tensors.

A typical vector in T_r^0 is a linear combination of **decomposable vectors** of type $\alpha \otimes \cdots \otimes \beta$. Indeed the set $\{\alpha^{i_1} \otimes \cdots \otimes \alpha^{i_r}\}, i_1, \ldots, i_r = 1, \ldots, n$ forms a basis in the n^r dimensional space T_r^0. Such tensors t are fully specified by n^r numbers $t_{i \ldots j} \equiv t(\mathbf{e}_i, \ldots, \mathbf{e}_j)$. A basis in this space is given by $\alpha^{i_1} \otimes \cdots \otimes \alpha^{i_r}$ with $i_1, \ldots, i_r = 1, \ldots, n$.

§ 19 : Spaces T_0^s

In a similar manner we can define space $T_0^s = V \otimes \cdots \otimes V$ of **contravariant tensors of rank** s as the set of all multilinear functionals on the cartesian product $V^* \times \cdots \times V^*$ (s-factors) with basis $\mathbf{e}_{i_1} \otimes \cdots \otimes \mathbf{e}_{i_s}$ with $i_1, \ldots, i_s = 1, \ldots, n$.

We can see that T_r^0 is dual to T_0^r.

Let $t = \beta \otimes \cdots \otimes \beta_r \in T_r^0 = V^* \otimes \cdots \otimes V^*$. Then t can be defined as a *linear* functional on $V \otimes \cdots \otimes V$ as follows : on decpomposable vectors it is

$$\begin{aligned} t(\mathbf{v}_1 \otimes \cdots \otimes \mathbf{v}_r) &= (\beta_1 \otimes \cdots \otimes \beta_r)(\mathbf{v}_1 \otimes \cdots \otimes \mathbf{v}_r) \\ &= \beta_1(\mathbf{v}_1) \cdots \beta_r(\mathbf{v}_r) \end{aligned}$$

while on vectors which are sums of these the linearity of the functional is used. This shows that t is in the space dual to T_0^r. Vector such as t span T_r^0 therefore $T_r^0 = (T_0^r)^*$.

§ 20 : Mixed tensor space T_1^1

Let T be a linear operator $T : V \to V$ on V. Then we can identify T with a bilinear functional on $V^* \times V$ (denoted by the same symbol T) by defining

$$T(\alpha, \mathbf{v}) = \alpha(T(\mathbf{v}))$$

It is easy to see that T indeed is bilinear. Therefore we can regard T as belonging to the tensor space $T^1_1 = V \otimes V^*$.

Just as we defined $\alpha \otimes \beta$ in $V^* \otimes V^*$, we can define, similarly $\mathbf{u} \otimes \beta \in V \otimes V^*$. It is a bilinear functional on $V^* \times V$ given by

$$(\mathbf{u} \otimes \beta)(\alpha, \mathbf{v}) = \alpha(\mathbf{u})\beta(\mathbf{v})$$

Of course, we can also interpret $\mathbf{u} \otimes \beta$ as a *linear* operator on V by defining $(\mathbf{u} \otimes \beta)(\mathbf{v}) = \beta(\mathbf{v})\mathbf{u}$

Members of the vector space$T^1_1 = V \otimes V^*$ are called mixed tensors of contravariant rank 1 and covariant rank 1. A basis can be chosen in this n^2 dimensional space by choosing $\{\mathbf{e}_i \otimes \alpha^j\}$ with i and j taking values $1, \ldots, n$.

To summarise, tensors of type T^1_1 can be considered as bilinear functionals on $V^* \times V$ or as linear mappings $V \to V$.

The space T^s_r of **mixed tensors** of contravariant rank s and covariant rank r (briefly called (s,r) tensors) can be defined as the set of multilinear maps on $V^* \times \cdots \times V^* \times V \times \cdots \times V$ (there are s factors of V^* and r of V). A typical multilinear map of this type is $\mathbf{v} \otimes \cdots \otimes \mathbf{w} \otimes \alpha \otimes \cdots \otimes \beta$ which gives acting on $V^* \times \cdots \times V^* \times V \times \cdots \times V$ as

$$(\mathbf{v} \otimes \cdots \otimes \mathbf{w} \otimes \alpha \otimes \cdots \otimes \beta)(\gamma, \ldots, \zeta, \mathbf{u}, \ldots, \mathbf{x}) = \gamma(\mathbf{v}) \cdots \zeta(\mathbf{w})\alpha(\mathbf{u}) \cdots \beta(\mathbf{x})$$

Such decomposable tensors form the basis

$$\mathbf{e}_{i_1} \otimes \cdots \otimes \mathbf{e}_{i_s} \otimes \alpha^{j_1} \otimes \cdots \otimes \alpha^{j_r}$$

A general multilinear map t in T^s_r is determines by its values

$$t^{i_1 \ldots i_s}{}_{j_1 \ldots j_r} \equiv t(\alpha^{i_1}, \ldots, \alpha^{i_s}, \mathbf{e}_{j_1}, \ldots, \mathbf{e}_{j_r})$$

which allows us to write

$$t = t^{i_1 \ldots i_s}{}_{j_1 \ldots j_r} \mathbf{e}_{i_1} \otimes \cdots \otimes \mathbf{e}_{i_s} \otimes \alpha^{j_1} \otimes \cdots \otimes \alpha^{j_r}$$

§ 21 : Interior product or contraction

Given a covariant tensor $t \in T^0_r$ of rank r and a vector $\mathbf{v}$ we define a tensor $i_{\mathbf{v}}t \in T^0_{r-1}$ of rank $r-1$ as follows :

$$(i_{\mathbf{v}}t)(\mathbf{v}_1, \ldots, \mathbf{v}_{r-1}) = t(\mathbf{v}, \mathbf{v}_1, \ldots, \mathbf{v}_{r-1}) \tag{14}$$

The linear mapping $i_{\mathbf{v}}$ satisfies the following properties :

$$\begin{aligned}
i_{\mathbf{v}}(t+s) &= i_{\mathbf{v}}(t) + i_{\mathbf{v}}(s) && (15)\\
i_{\mathbf{v}}(at) &= a i_{\mathbf{v}}(t) && (16)\\
i_{\mathbf{v}+\mathbf{u}}(t) &= i_{\mathbf{v}}(t) + i_{\mathbf{u}}(t) && (17)\\
i_{a\mathbf{v}}(t) &= a i_{\mathbf{v}}(t) && (18)
\end{aligned}$$

Similarly, if we are given a mixed (r,s)-tensor T with contra- and covariant indices then a **contraction** between the k-th contra- and l-th covariant index is defined as an $(r-1, s-1)$-tensor $i(k,l)T$ as follows

$$\begin{aligned}
&(i(k,l)T)(\beta^1, \beta^2, \ldots, \beta^{r-1}, \mathbf{v}_1, \ldots, \mathbf{v}_{s-1})\\
&= \sum_i T(\beta^1, \beta^2, \ldots, \alpha^i, \ldots, \beta^{r-1}, \mathbf{v}_1, \ldots, \mathbf{e}_i, \ldots, \mathbf{v}_{s-1}) \qquad (19)
\end{aligned}$$

where the dual basis elements α^i and $\mathbf{e}_i$ appear on the k-th and l-th place. Simply told, in terms of components,

$$\begin{aligned}
&(i(k,l)T)^{i_1 \ldots i_{k-1} i_{k+1} \ldots i_r}{}_{j_1 \ldots j_{l-1} j_{l+1} \ldots j_s}\\
&= T^{i_1 \ldots i_{k-1} j i_{k+1} \ldots i_r}{}_{j_1 \ldots j_{l-1} j j_{l+1} \ldots j_s} \qquad (20)
\end{aligned}$$

§ 22 : Summary

To summarize, we start with a vector space V with a basis $\{\mathbf{e}_i\}$ and define the dual space V^* with a dual basis $\{\alpha^i\}$. With these spaces and bases as starting point we can define an infinite sequence of vector spaces of higher and higher dimensions with tensor products. The components of a tensor are characterised by the way they transform when we change the basis $\{\mathbf{e}_i\}$ to a new basis. This leads to a change in the dual basis, and to bases in all the tensor spaces.

As a matter of standard established notation observe carefully the use of super- and sub-scripts for denoting the members of bases in V^* and V (respectively) as well as in the components of contra- and co-variant vectors and tensors.

We can verify the transformation properties of components of the covariant and contravariant tensors : when the basis $\{\mathbf{e}_i\}$ in V is changed to $\{\mathbf{f}_i\}$ as

$$\mathbf{f}_i = T_i^{\ j}\mathbf{e}_i$$

the dual basis changes from $\{\alpha^i\}$ to $\{\beta^i\}$

$$\beta^i = (T^{-1T})^i{}_k \alpha^k$$

and components of tensors change as

$$t'_{i\ldots j} = T_i^{\ k} \ldots T_j^{\ l} t_{k\ldots l} \tag{21}$$

$$t'^{i\ldots j} = (T^{-1T})^i{}_k \ldots (T^{-1T})^j{}_l t^{kl} \tag{22}$$

$$t'^{i\ldots j}{}_{k\ldots l} = (T^{-1T})^i{}_p \ldots (T^{-1T})^j{}_q T_k^{\ m} \ldots T_l^{\ n} t^{p\ldots q}{}_{m\ldots n} \tag{23}$$

where the primes denote components with respect to the bases β^i and $\mathbf{f}_i$.

§ 23 : Wedge or Exterior Product

Antisymmetric covariant tensors are extremely important because of their connection to differential forms, surface and volume integrals and Gauss-Stokes theorem. This importance is reflected in the fact that the antisymmetric part of tensor products have a different symbol and name to denote it.

For second rank tensors we write

$$\alpha \wedge \beta = \alpha \otimes \beta - \beta \otimes \alpha$$

and generalise it to tensors of higher rank.

§ 24 : Permutations

Let P be a permutation of r objects. This means that P is a one-to-one mapping of the set $\{1, \ldots, r\}$ of first r natural numbers onto itself. There are $r!$ such mappings. Each of these can be considered as composed of more elementary mappings called transpositions, which just exchange two of the integers and map the rest to themselves. The way in which transpositions make a permutation P is not unique but the number of transpositions involved though not fixed is either always an even or an odd integer. The permutation is called even or odd accordingly. Let us define $(-1)^P$ to be equal to $+1$ if P is even and -1 if odd. The set of $r!$ permutations can be made a group under composition of mappings as the group law. The identity mapping is even. It is also clear that $(-1)^{P_1 \circ P_2} = (-1)^{P_1}(-1)^{P_2}$. It follows that P^{-1} is even or odd according as P.

Let us define a linear operator on T_r^0 corresponding to permutation P, also to be denoted by P, as follows. Let $\beta_1, \ldots, \beta_r$ be vectors in V^*. Form the tensor product $\beta_1 \otimes \cdots \otimes \beta_r$. Now define

$$P(\beta_1 \otimes \cdots \otimes \beta_r) \equiv \beta_{P(1)} \otimes \cdots \otimes \beta_{P(r)} \tag{24}$$

It is sufficient to define P on these vectors because any general vector in T_r^0 is a linear combination of such vectors, and P

is linear. For two permutations P_1 and P_2 we have clearly $P_2 P_1 = P_2 \circ P_1$ where in this equation the linear operators are on the left and the mappings on the right.

§ 25 : Exterior or wedge product

We are now ready to define the wedge product of any number of vectors of V^*.

$$\begin{aligned}\beta_1 \wedge \beta_2 \wedge \cdots \wedge \beta_r &\equiv \sum_P (-1)^P P(\beta_1 \otimes \cdots \otimes \beta_r) \\ &= \sum_P (-1)^P \beta_{P(1)} \otimes \cdots \otimes \beta_{P(r)} \qquad (25)\end{aligned}$$

If Q is a permutation operator, then

$$\begin{aligned}\beta_{Q(1)} \wedge \beta_{Q(2)} \wedge \cdots \wedge \beta_{Q(r)} &= \sum_P (-1)^P PQ(\beta_1 \otimes \cdots \otimes \beta_r) \\ &= (-1)^Q \sum_R (-1)^R R(\beta_1 \otimes \cdots \otimes \beta_r)\end{aligned}$$

where we have used $R \equiv P \circ Q$ and $(-1)^R = (-1)^P(-1)^Q = (-1)^P/(-1)^Q$. Therefore,

$$\beta_{Q(1)} \wedge \beta_{Q(2)} \wedge \cdots \wedge \beta_{Q(r)} = (-1)^Q(\beta_1 \wedge \beta_2 \wedge \cdots \wedge \beta_r) \qquad (26)$$

In particular, the wedge product like $\beta_1 \wedge \beta_2 \wedge \cdots \wedge \beta_r$ changes sign whenever any two factors in it are exchanged. Therefore if any factor is repeated, the product is the zero vector.

Tensors with this property are called **antisymmetric**. As a multilinear functional on $V \times \cdots \times V$,

$$(\beta_1 \wedge \beta_2 \wedge \cdots \wedge \beta_r)(\mathbf{v}_1, \ldots, \mathbf{v}_r) = \det \|\beta_i(\mathbf{v}_j)\| \qquad (27)$$

Linear combinations of antisymmetric tensors are also antisymmetric, therefore the set of all antisymmetric covariant tensors forms a subspace $\Lambda^r(V^*) \subset T^0_r = V^* \otimes \cdots \otimes V^*$. These tensors are also called r-**forms**.

§ 26 : Bases for $\Lambda^r(V^*)$

A basis in $\Lambda^r(V^*)$ can be chosen by considering all independent tensors of the form $\alpha^{i_1} \wedge \cdots \wedge \alpha^{i_r}$. Obviously, $i_1, \ldots, i_r$ all have to be different, because the antisymmetric wedge product is zero if any two vectors in the product string are the same. Also, a particular combination $i_1, \ldots, i_r$ need be taken anly once because any other product with these same indices (though in some other order) is ± 1 times the same vector. There are as many independent tensors of this type as the number of ways to choose a combination of r different indices $i_1, \ldots, i_r$ out of $1, \ldots, n$. The dimension of space $\Lambda^r(V^*)$ is therefore $n!/r!(n-r)!$.

A basis can be chosen consisting of vectors $\{\alpha^{i_1} \wedge \cdots \wedge \alpha^{i_r}\}$ with $i_1 < \cdots < i_r$.

For $r = n$ the space $\Lambda^n(V^*)$ is one dimensional containing multiples of $\alpha^1 \wedge \cdots \wedge \alpha^n$. For $r > n$ the spaces Λ^r are zero, that is, contain only the zero vector.

§ 27 : Space $\Lambda^r(V)$

In exactly the same manner we define the space $\Lambda^r(V)$ of all antisymmetric contravariant tensors. They are called r-**vectors**. This space is spanned by $\{\mathbf{e}_{i_1} \wedge \cdots \wedge \mathbf{e}_{i_r}\}$ with $i_1 < \cdots < i_r$.

§ 28 : Wedge product of an r- and an s-form

Given an r-form $t \in \Lambda^r(V^*)$ and an s-form $u \in \Lambda^s(V^*)$ we can define an $r+s$-form $t \wedge u$ called the wedge or exterior product of t and u as follows. First define it on decomposable vectors :

$$(\beta_1 \wedge \cdots \wedge \beta_r) \wedge (\gamma_1 \wedge \cdots \wedge \gamma_s) = \beta_1 \wedge \cdots \wedge \beta_r \wedge \gamma_1 \wedge \cdots \wedge \gamma_s$$

and then extend it on general vectors by linearity. Because

$$\beta_1 \wedge \cdots \wedge \beta_r \wedge \gamma_1 \wedge \cdots \wedge \gamma_s = (-1)^{rs} \gamma_1 \wedge \cdots \wedge \gamma_s \wedge \beta_1 \wedge \cdots \wedge \beta_r$$

follows from the antisymmetry of the wedge product, we must have in general

$$t \wedge u = (-1)^{rs} u \wedge t \tag{28}$$

§ 29 : Bases in T_r^0 and $\Lambda^r(V^*)$

Let $t \in \Lambda^r(V^*)$ be written

$$t = \sum_{i_1 < \cdots < i_r} T_{i_1 \ldots i_r} \alpha^{i_1} \wedge \cdots \wedge \alpha^{i_r}$$

Note that coefficients $T_{i_1 \ldots i_r}$ are defined only for indices $i_1 < \cdots < i_r$.

As $\Lambda^r(V^*) \subset T_r^0$, the r-form t can also be expanded as a member of T_r^0 in the basis $\{\alpha^{j_1} \otimes \cdots \otimes \alpha^{j_r}\}_{j_1, \ldots, j_r = 1}^n$, with coefficients $t_{j_1, \ldots, j_r} = t(\mathbf{e}_{j_1}, \ldots, \mathbf{e}_{j_r})$

$$t = \sum_{j_1, \ldots, j_r = 1}^{n} t_{j_1, \ldots, j_r} \alpha^{j_1} \otimes \cdots \otimes \alpha^{j_r}$$

By expanding the wedge product basis vectors in terms of tensor product basis, and then comparing the coefficients we find the components t's in terms of T's.

$$t_{j_1 \ldots j_r} = 0 \text{ if any of the indices coincide}$$

$$t_{j_1 \dots j_r} = T_{j_1 \dots j_r} \qquad \text{for } j_1 < \cdots < j_r$$

$$t_{j_1 \dots j_r} = (-1)^P T_{P(j_1) \dots P(j_r)}$$

where permutation P brings indices $j_1 \dots j_r$ to increasing order $P(j_1) < \cdots < P(j_r)$

§ 30 : Components of $t \wedge u$ in $\Lambda^{r+s}(V^*)$ basis

Given that

$$t = \sum_{i_1 < \cdots < i_r} T_{i_1 \dots i_r} \alpha^{i_1} \wedge \cdots \wedge \alpha^{i_r}$$

$$u = \sum_{i_1 < \cdots < i_s} U_{i_1 \dots i_s} \alpha^{i_1} \wedge \cdots \wedge \alpha^{i_s}$$

we can work out the components $B_{i_1 \dots i_{r+s}}$ of $t \wedge u$ in the basis $\{\alpha^{i_1} \wedge \cdots \wedge \alpha^{i_{r+s}}\}$. They are

$$B_{i_1 \dots i_{r+s}} = \sum_{(r,s) \text{ shuffles } \mathrm{Q}} T_{Q(i_1) \dots Q(i_r)} U_{Q(i_{r+1}) \dots Q(i_{r+s})} \tag{29}$$

where the sum is over all (r,s) shuffles defined below.

An (r,s) **shuffle** is defined to be a permutation Q of $(r+s)$ distinct integers $(i_1 < \cdots < i_{r+s})$ such that

$$[i_1, \cdots, i_{r+s}] \to [Q(i_1), \cdots, Q(i_r) \ ; \ Q(i_{r+1}), \cdots Q(i_{r+s})]$$

where $Q(i_1) < \cdots < Q(i_r)$ and $Q(i_{r+1}) < \cdots < Q(i_{r+s})$.

The total number of (r,s) shuffles is $(r+s)!/r!s!$.

As an example (2,4,5,7) have the following six (2,2) shuffles :

$$(2,4,5,7) \to (2,4;5,7), (2,5;4,7), (2,7;4,5), (4,5;2,7), (4,7;2,5), (5,7;2,4)$$

§ 31 : Components of $t \wedge u$ in T^0_{r+s} basis

Given an r-form t and an s-form u whose antisymmetric components as members of T^0_r and T^0_s respectively are

$$t = t_{i_1,\ldots,i_r} \alpha^{i_1} \otimes \cdots \otimes \alpha^{i_r}$$

and

$$u = u_{j_1,\ldots,j_s} \alpha^{j_1} \otimes \cdots \otimes \alpha^{j_s}$$

we can show that the components $p_{k_1,\ldots,k_{r+s}}$ of $p = t \wedge u$

$$p = t \wedge u = p_{k_1,\ldots,k_{r+s}} \alpha^{k_1} \otimes \cdots \otimes \alpha^{k_{r+s}}$$

are given by

$$p_{k_1,\ldots,k_{r+s}} = (1/r!s!) \sum_P (-1)^P t_{P(k_1),\ldots,P(k_r)} u_{P(k_{r+1}),\ldots,P(k_{r+s})}$$

LECTURE 3 : INNER PRODUCT OR METRIC

A symmetric bilinear non-degenerate form on a vector space is called a metric or inner product. It helps us to define the notion of orthogonality. Given a metric one can construct orthonormal bases, which are very useful. Moreover, the metric sets up a one-to-one correspondence between the vector space and its dual, called 'raising and lowering the indices' by physicists. Components of the metric in tangent spaces of spacetime play a central role in general relativity as the gravitational field.

§ 32 : Vector space

A vector space is defined by the operations of sum of its vectors and multiplication by real numbers to its vectors.

An inner product or metric is an additional structure on a vector space.

For any two vectors $\mathbf{v}$ and $\mathbf{w}$ in a vector space V their **inner product** is a real number denoted by $\langle \mathbf{v}, \mathbf{w} \rangle$. The function which defines the inner product should have the following properties:

1. it is **linear**, that is,

$$\langle \mathbf{u}, \mathbf{v} + \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{u}, \mathbf{w} \rangle \tag{30}$$

$$\langle \mathbf{u}, a\mathbf{v} \rangle = a\langle \mathbf{u}, \mathbf{v} \rangle \tag{31}$$

   for any $\mathbf{v}, \mathbf{u}, \mathbf{w} \in V$ and any real number a.

2. it is **symmetric**

$$\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle \tag{32}$$

   for any $\mathbf{v}, \mathbf{u} \in V$ With this property we can see that the inner product is linear in the the first argument as well :

$$\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle$$

$$\langle a\mathbf{u}, \mathbf{v} \rangle = a\langle \mathbf{u}, \mathbf{v} \rangle$$

3. and it is **non-degenerate**, that is, if $\langle \mathbf{u}, \mathbf{v} \rangle = 0$ for all $\mathbf{v} \in V$ then $\mathbf{u} = \mathbf{0}$.

The inner product is also called the **metric**.

An inner product is often defined with a stronger condition of **positive definiteness** which says that $\langle \mathbf{v}, \mathbf{v} \rangle \geq 0$ and $\langle \mathbf{v}, \mathbf{v} \rangle = 0$ if and only if $\mathbf{v} = \mathbf{0}$.

Obviously, a positive definite inner product is non-degenerate (take $\mathbf{v} = \mathbf{u}$ in the condition of non-degeneracy), but not vice versa.

In relativity theory we need inner-products on spacetime which are not positive definite. In Minkowski space there are time-like vectors whose inner-product with themselves is negative or null vectors for which it is zero. But the inner product is always non-degenerate.

Given an inner product we can define the notion of **orthogonality**. Two vectors $\mathbf{v}$ and $\mathbf{u}$ in V are called **orthogonal** if their inner product is zero, that is $\langle \mathbf{v}, \mathbf{u} \rangle = 0$.

For any vector $\mathbf{v}$ the number $\langle \mathbf{v}, \mathbf{v} \rangle$ is its **norm squared**. When the inner product is positive definite, as in Euclidean space, the positive number $\sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$ is its **norm** or **length**.

When the inner product is not positive definite, norm squared can be positive, negative or zero. A vector with zero norm squared (that is a vector which is orthogonal to itself) is called a **null** vector. A vector with norm squared equal to ± 1 is called **normalized**.

Two non-null, orthogonal vectors are linearly independent. What is interesting is that two non-orthogonal null vectors are also linearly independent. And in (3+1)dimensional Minkowski space, two null vectors if they are orthogonal, then they are necessarily proportional to each other.

§ 33 : Orthonormal Bases

An inner product or metric as defined above is a bilinear functional on $V \times V$. Therefore it defines a second rank, symmetric covariant tensor $\mathbf{g}$ called the **metric tensor** through

$$\mathbf{g}(\mathbf{u}, \mathbf{v}) = \langle \mathbf{u}, \mathbf{v} \rangle$$

We could equally well use the notation with $\mathbf{g}(\mathbf{u}, \mathbf{v})$ in place of $\langle \mathbf{u}, \mathbf{v} \rangle$. In most cases however, $\mathbf{g}$ is a given, fixed tensor and there is ease of notation in using the bracket notation for the inner-product. In three dimensional vector spaces the notation used is $\mathbf{u} \cdot \mathbf{v}$ instead of $\langle \mathbf{u}, \mathbf{v} \rangle$.

Let $\mathbf{e}_i, i = 1, \ldots, n$ be a basis in V. The components of the metric in this basis are

$$g_{ij} = \langle \mathbf{e}_i, \mathbf{e}_j \rangle \tag{33}$$

This symmetric matrix contains all the information about the inner product because if $\mathbf{v} = v^i \mathbf{e}_i$ and $\mathbf{u} = u^j \mathbf{e}_j$ are two vectors then bilinearitry of the product in its two arguments implies

$$\langle \mathbf{v}, \mathbf{u} \rangle = g_{ij} v^i u^j \tag{34}$$

The non-degeneracy of the inner product means that $g \equiv \det g_{ij} \neq 0$ or, in other words, the matrix of metric tensor components in any basis is non-singular.

We note the important result that despite the existence of null (zero-norm) vectors in the space one can always choose a basis $\{\mathbf{n}_i\}_{i=1}^n$ such that $\eta_{ij} = \langle \mathbf{n}_i, \mathbf{n}_j \rangle = 0$ if $i \neq j$ and the norm squared $\langle \mathbf{n}_i, \mathbf{n}_i \rangle$ is either $+1$ or -1. Such a basis is called an **orthonormal basis**. We construct such a basis in the next section.

The number of vectors with norm squared $+1$ and those with norm squared -1 in such an orthonormal basis is fixed by the definition of the inner product. The number of positive norm squared vectors minus the number of negative norm squared vectors in an orthonormal basis is called the **signature** of the metric and is denoted by $\mathrm{sig}(V)$.

There do exist bases which contain (non-zero) null vectors as basis vectors. But these bases are not orthonormal.

§ 34 : Existence of orthonormal bases

We go through the standard proof of the existence of orthonormal bases because of its fundamental importance.

Let $\mathbf{a}$ be a non-zero vector with non-zero norm squared $\langle \mathbf{a}, \mathbf{a} \rangle \neq 0$.

There certainly exists a non-null vector of this kind unless the whole space is trivial consisting of just the zero vector $\mathbf{0}$. This is so because if $\langle \mathbf{a}, \mathbf{a} \rangle$ were zero for all $\mathbf{a} \in V$ then by using $\langle \mathbf{a}+\mathbf{b}, \mathbf{a}+\mathbf{b} \rangle = 0$ for any arbitrary $\mathbf{b}$ it follows that $\langle \mathbf{a}, \mathbf{b} \rangle = 0$ for all $\mathbf{b}$. The condition of non-degeneracy then implies $\mathbf{a} = \mathbf{0}$.

Let $\mathbf{n}_1 = \mathbf{a}/\sqrt{|\langle \mathbf{a}, \mathbf{a} \rangle|}$. Depending on the sign of norm squared of $\mathbf{a}$, $\langle \mathbf{n}_1, \mathbf{n}_1 \rangle \equiv \epsilon_1$ is $+1$ or -1.

Let V_1 be the one dimensional subspace spanned by $\mathbf{n}_1$, and let V_2 be the set of all vectors in V orthogonal to every vector in V_1. Obviously, V_2 is a vector subspace, and every vector $\mathbf{v} \in V$ can be decomposed as

$$\mathbf{v} = \epsilon_1 \langle \mathbf{v}, \mathbf{n}_1 \rangle \mathbf{n}_1 + (\mathbf{v} - \epsilon_1 \langle \mathbf{v}, \mathbf{n}_1 \rangle \mathbf{n}_1)$$

where the first term is in V_1 and the second term is in V_2. The only vector common to V_1 and V_2 is the zero vector $\mathbf{0}$ and this again follows from non-degeneracy.

Therefore $V = V_1 \oplus V_2$ and the inner product restricted to V_2 is again non-degenerate because a vector in V_2 orthogonal to all other vectors of V_2 is moreover orthogonal to V_1 and hence is zero-vector.

We can now start with V_2 as the starting space and find a non null vector $\mathbf{b} \in V_2$ such that $\langle \mathbf{b}, \mathbf{b} \rangle \neq 0$, and construct $\mathbf{n}_2 = \mathbf{b}/\mathbf{b}/\sqrt{|\langle \mathbf{b}, \mathbf{b} \rangle|}$, with $\langle \mathbf{n}_2, \mathbf{n}_2 \rangle \equiv \epsilon_2$ equal to +1 or -1. We proceed in this manner inductively till the whole basis is constructed.

Thus we have a basis $\{\mathbf{n}_i\}$ with the metric components

$$I_{\epsilon ij} = \langle \mathbf{n}_i, \mathbf{n}_j \rangle = \epsilon_i \delta_{ij} \qquad \text{(no summation on } i\text{)} \tag{35}$$

or

$$I_\epsilon = \begin{pmatrix} \epsilon_1 & 0 & \dots & 0 \\ 0 & \epsilon_2 & \dots & 0 \\ & & \ddots & \\ 0 & 0 & \dots & \epsilon_n \end{pmatrix} \tag{36}$$

§ 35 : Signature of the metric

Note that whichever route we take to choose orthonormal vectors for a basis the number n_+ of vectors with norm $+1$ and the number n_- of vectors with norm -1 is always the same. As $\dim V = n = n_+ + n_-$ is fixed so is the number $t = n_+ - n_-$. t is called the **signature** of the metric. The Minkowski space has one timelike unit vector with $\epsilon_0 = -1$ and three spacelike vectors with $(\epsilon_i = 1, i = 1, 2, 3)$ in any orthonormal basis. Thus it has signature $+2$.

Similarly the number $\pm 1 = \epsilon_1 \epsilon_2 \ldots \epsilon_n = \det I_\epsilon$ which the determinant of the matrix I_ϵ of metric components (in the orthonormal basis) is a characteristic of the metric. If $\{\mathbf{e}_j\}$ is any basis with $g_{ij} = \langle \mathbf{e}_i, \mathbf{e}_j \rangle$ as the metric components then $g \equiv \det \|g_{ij}\|$ (which is always non-zero) has the same sign as $\det I_\epsilon$. We write this number as $\operatorname{sgn}(g)$ in general.

For spacetime in general relativity the sign of $g = \det g_{ij}$ is always negative because the number of vectors with negative norm is odd.

§ 36 : Correspondence between V and V^*

A non-degenerate inner product $\langle \; , \; \rangle$ defined on a vector space V *sets up a one-to-one correspondence* between vectors in V and those in the dual V^*.

Let $\mathbf{v} \in V$ be given. Then every vector $\mathbf{u} \in V$ can be mapped linearly to real numbers by $\mathbf{u} \to \langle \mathbf{v}, \mathbf{u} \rangle$. This helps us define a linear functional $\mathbf{v}^\flat \in V^*$ (called "$\mathbf{v}$-flat") with the help of the inner product as

$$\mathbf{v}^\flat(\mathbf{u}) = \langle \mathbf{v}, \mathbf{u} \rangle$$

Properties of the inner product ensure that $\mathbf{v}^\flat$ is a linear functional. Obviously, $\mathbf{v}^\flat$ depends on the vector $\mathbf{v}$.

In fact, for finite dimensional spaces, *all* linear functionals arise in this way. In other words ,if we are given a linear functional $\alpha \in V^*$ then there exists a vector $\alpha^\sharp \in V$ ("alpha-sharp") such that the number assigned by α to a vector $\mathbf{u}$ is the same number $\langle \alpha^\sharp, \mathbf{u} \rangle$ obtained in taking inner product with $\alpha^\sharp$.

We can identify $\alpha^\sharp$ as follows.

First choose an orthonormal basis $\{\mathbf{n}_i\}$ with $\langle \mathbf{n}_i, \mathbf{n}_j \rangle = 0$ for $i \neq j$ and $\langle \mathbf{n}_i, \mathbf{n}_i \rangle = \epsilon_i$.

Now define

$$\alpha^\sharp = \sum_i \epsilon_i \alpha(\mathbf{n}_i) \mathbf{n}_i \tag{37}$$

This is indeed the vector with the required property. Expanding $\mathbf{u} = u^j \mathbf{e}_j$ and using $(\epsilon_i)^2 = 1$

$$\begin{aligned} \langle \alpha^\sharp, \mathbf{u} \rangle &= \sum_j u^j \langle \alpha^\sharp, \mathbf{n}_j \rangle = \sum_{ij} u^j \epsilon_i \alpha(\mathbf{n}_i) \langle \mathbf{n}_i, \mathbf{n}_j \rangle \\ &= \sum_{ij} u^j \epsilon_i \alpha(\mathbf{n}_i) \epsilon_i \delta_{ij} = \sum_j u^j \alpha(\mathbf{n}_j) = \alpha(\mathbf{u}) \end{aligned}$$

This one-to-one correspondence between the the dual spaces V and V^* is called **raising and lowering of indices** by the matrix g_{ij}. The reason for this nomenclature is as follows.

If $\mathbf{v} \in V$ has components v^i in basis $\mathbf{e}_i$ (not necessarily orthonormal) then the linear functional $\mathbf{v}^\flat \in V^*$ corresponding to it has components

$$a_i = g_{ij} v^j$$

with respect to the dual basis α^i.

$$\mathbf{v}^\flat(\mathbf{u}) =< \mathbf{v}, \mathbf{u} >= g_{ij} v^i u^j = g_{ij} v^i \alpha^j(\mathbf{u}) \equiv (a_j \alpha^j)(\mathbf{u}) \tag{38}$$

where we have used the property of dual basis : if $\mathbf{u} = u^j \mathbf{e}_j$ then $\alpha^i(\mathbf{u}) = u^i$.

The inverse of this one-to-one correspondence between components of a linear functional $\alpha = a_i \alpha^i$ to a vector $\alpha^\sharp = v^j \mathbf{e}_j$ is similarly given by

$$v^i = g^{ij} a_j$$

where g^{ij} (with matrix indices written as superscripts) is the *inverse* of the matrix g_{ij}. We see this in the next section.

This also explains the musical notation of sharp and flat. A vector $\mathbf{v} \in V$ (contra variant,upper index) corresponds, via the metric, to a form $\mathbf{v}^\flat$ (covariant, lower index=lower pitch=flat), while a form α becomes a vector (contravariant=upper index=higher pitch=sharp) $\alpha^\sharp \in V^*$.

§ 37 : Remark on physicists' notation

Most general relativity physicists use the same letter-symbol for components when a (contrvariant) vector $\mathbf{v}$ is put in correspondence with a (covariant) form $\mathbf{v}^\flat$ or vice versa. For example if

$$\mathbf{v} = v^i \mathbf{e}_i$$

then $\mathbf{v}^\flat$ is written as

$$\mathbf{v}^\flat = v_i \alpha^i$$

in the dual basis. The correspondence itself looks like

$$v_i = g_{ij} v^j, \qquad v^i = g^{ij} v_j$$

This notation although convenient can be very confusing in geometrical contexts. For example, in dealing with a hypersurface, the normal vector field on the surface and its corresponding 1-form have very different roles to play, and the choice of using same symbols for components is more a liability than convenience.

§ 38 : Inner Product in V^*

An inner product defined on a vector space V, determines in a natural way an inner product on the dual vector space V^* as well as on spaces T^r_s and $\Lambda^r(V^*)$ etc.

We saw in the last section that to every $\alpha \in V^*$ corresponds a vector $\alpha^\sharp$ such that $\alpha(\mathbf{u}) = \langle \alpha^\sharp, \mathbf{u}\rangle$ for every $\mathbf{u} \in V$ and conversely every vector in V determines a member of V^* in this manner.

This one-to-one correspondence suggests that if α corresponds to $\alpha^\sharp = \sum_i \epsilon_i \alpha(\mathbf{n}_i)\mathbf{n}_i$ for an orthonormal basis $\{\mathbf{n}_i\}$ and similarly if β corresponds to $\beta^\sharp$, then we can define

$$\langle \alpha, \beta\rangle \equiv \langle \alpha^\sharp, \beta^\sharp\rangle = \sum_i \epsilon_i \alpha(\mathbf{n}_i)\beta(\mathbf{n}_i)$$

All properties of the inner-product are satisfied. Perhaps the only property not obvious to see is non-degeneracy. We follow it up in an exercise.

In a basis with vectors $\mathbf{e}_i = {T_i}^k \mathbf{n}_k$ the metric is

$$g_{ij} = \langle \mathbf{e}_i, \mathbf{e}_j\rangle = {T_i}^k \epsilon_k {T_j}^k = (T I_\epsilon T^T)_{ij}$$

where I_ϵ is the matrix of the metric in the orthonormal basis

$$(I_\epsilon)_{ij} = \langle \mathbf{n_i}, \mathbf{n}_j\rangle = \epsilon_i \delta_{ij} \qquad \text{no summation} \tag{39}$$

The convenience of the matrix in the orthonormal basis is that its square is the identity marix and $(I_\epsilon)^{-1} = (I_\epsilon)$.

Let $\{\alpha^i\}$ be the basis dual to $\{\mathbf{e}_i\}$. The dual basis is obtained by matrix T^{-1T} acting on the basis $\{\nu^i\}$ dual to $\{\mathbf{n}_i\}$

$$\alpha_i = \sum_l (T^{-1})_l{}^i \nu_l$$

The matrix of the metric in the dual space is then

$$\begin{aligned} g^{ij} &\equiv \langle \alpha^i, \alpha^j \rangle && (40) \\ &= \sum_k \epsilon_k \alpha^i(\mathbf{n}_k)\alpha^j(\mathbf{n}_k) \\ &= \sum_{k,l,m} (T^{-1})_l{}^i \nu_l(\mathbf{n}_k)\epsilon_k (T^{-1})_m{}^j \nu_m(\mathbf{n}_k) \\ &= (T^{-1T} I_\epsilon T^{-1})^{ij} \\ &= (g^{-1})_{ij} && (41) \end{aligned}$$

which follows from the relation $g_{ij} = (T I_\epsilon T^T)_{ij}$ given above.

Therefore we see that the matrix $g^{ij} = \langle \alpha^i, \alpha^j \rangle$ of the naturally determined metric in the dual basis of V^* is the inverse of the matrix $g_{ij} = \langle \mathbf{e}_i, \mathbf{e}_j \rangle$ in the original basis.

§ 39 : Inner product in tensor spaces

We now define the inner inner product on $T^r_0 = V \otimes \cdots \otimes V$.

For $\mathbf{a}_1 \otimes \mathbf{a}_2 \cdots \otimes \mathbf{a}_r \in T^r_0$ define the inner product on decomposable vectors as

$$\langle \mathbf{a}_1 \otimes \mathbf{a}_2 \cdots \otimes \mathbf{a}_r, \mathbf{b}_1 \otimes \mathbf{b}_2 \cdots \otimes \mathbf{b}_r \rangle \equiv \langle \mathbf{a}_1, \mathbf{b}_1 \rangle \cdots \langle \mathbf{a}_r, \mathbf{b}_r \rangle$$

and extend by linearity over arbitrary linear combinations.

To check the non-degeneracy property of this inner product let us choose an orthonormal basis $\{\mathbf{n}_i\}$ in V with $\langle \mathbf{n}_i, \mathbf{n}_j \rangle =$

$\pm\delta_{ij}$. Let $t = t^{i_1 \dots i_r} \mathbf{n}_{i_1} \otimes \cdots \otimes \mathbf{n}_{i_r}$ be such that $\langle t, \mathbf{n}_{j_1} \otimes \cdots \otimes \mathbf{n}_{j_r} \rangle = 0$ for all sets $\{j_1, \dots j_r\}$. Then it follows from the definition that $t^{i_1 \dots i_r} = 0$ for all $i_1 \dots i_r$, that is $t = \mathbf{0}$.

An exactly similar definition can be given for the inner product in $T^0_r(V) = V^* \otimes \cdots \otimes V^*$.

Tutorial 1 : Exercises and problems : Lectures 1 to 3

Exercise 1 Derive the 'velocity addition theorem'

$$V = \frac{v + v^*}{1 + vv^*/c^2}$$

from the group law of Lorentz transformations.

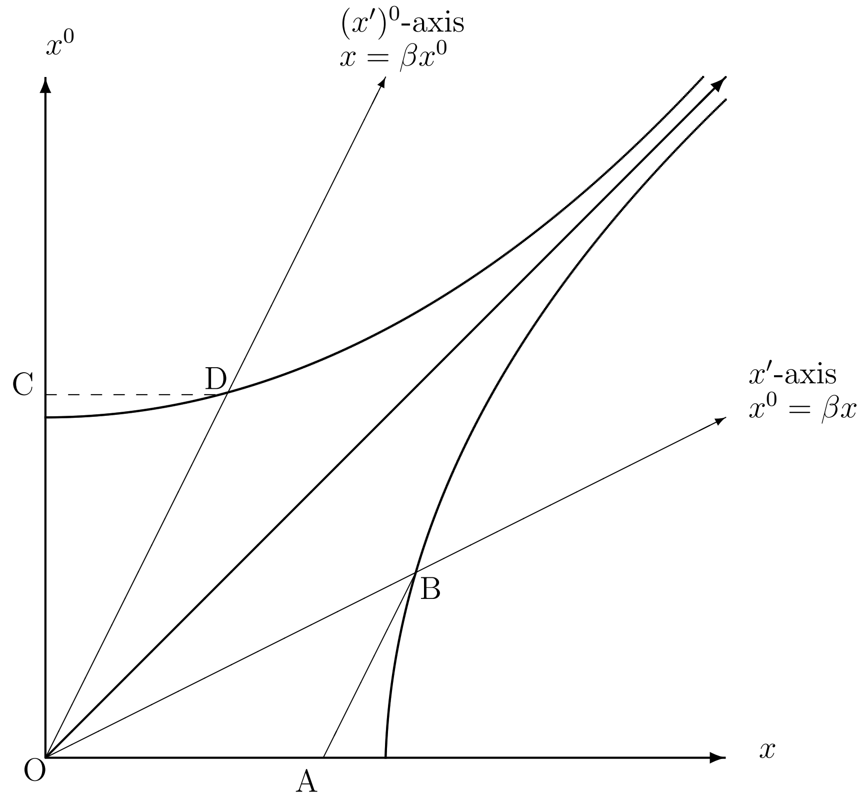
Exercise 2 Discuss the case when α is a large positive number.

Exercise 3 Show the phenomenon of Lorentz contraction and time dilatation on the Minkowski diagram.

Hint:

Figure out the diagram below. A rod of unit length is kept along the x'-axis with one end at the origin. Similarly, a clock is kept at the origin of frame S'. They both move with the frame S'. Check that the various points (or events) O, A, B, C, D have the following coordinates in the two frames.

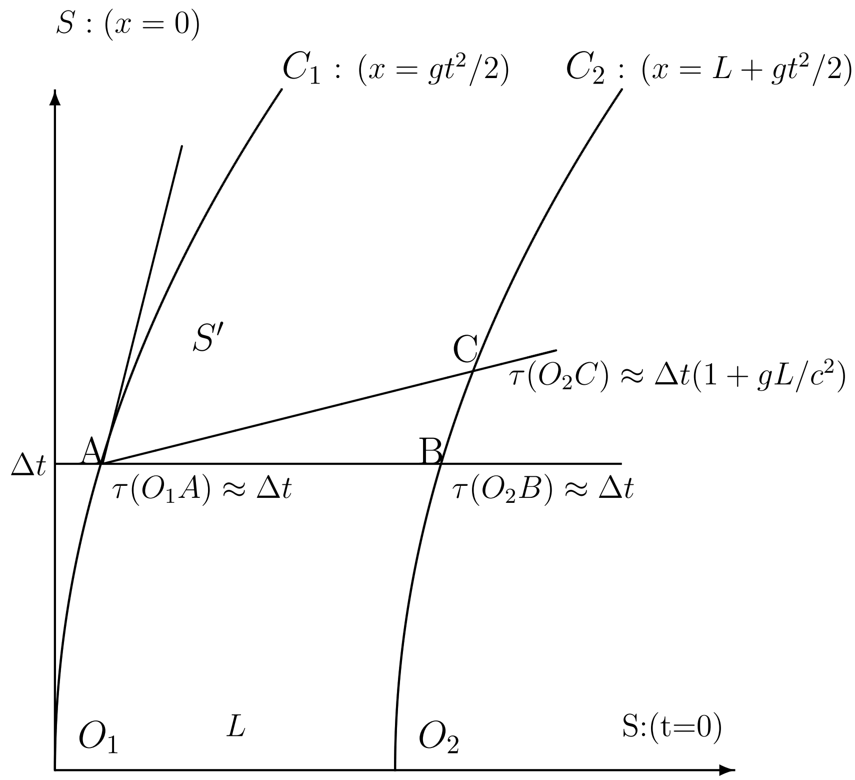
point	x	x^0	x'	$(x')^0$
O	0	0	0	0
A	$\sqrt{1-\beta^2}$	0	1	$-\beta/\sqrt{1-\beta^2}$
B	$1/\sqrt{1-\beta^2}$	$\beta/\sqrt{1-\beta^2}$	1	0
C	0	$1/\sqrt{1-\beta^2}$	$-\beta/(1-\beta^2)$	$1/(1-\beta^2)$
D	$\beta/\sqrt{1-\beta^2}$	$1/\sqrt{1-\beta^2}$	0	1



Exercise 4 Follow Einstein 1907 argument to show clocks run slowly in near gravitating body using equivalece principle.

Solution:

Let there be an inertial frame S with respect to which another frame S_1 is at rest at $t = 0$ with coinciding axes and the origin. Let S_1 start accelerating in the x-direction at $t = 0$ with acceleration g.



Einstein's argument of 1907

Clocks run slowly at lower gravitational potential.

Let there be two clocks C_1 and C_2 in the accelerating frame S_1. They are at events O_1 and O_2 at $t = 0$ as seen by S. As shown the trajectories of the clocks C_1 and C_2 according to frame S are given by $x_1(t)$ and $x_2(t)$ where

$$x_1(t) = \frac{1}{2}gt^2, \qquad x_2(t) = L + \frac{1}{2}gt^2$$

These equations describe the accelerated frame at low enough velocities, that is, at small values of t.

After a small time Δt by the clock of S, the frame S_1 is moving with velocity $g\Delta t$, and the clocks in S_1 are located respectively at events $A : (x = g(\Delta t)^2/2, t = \Delta t)$ and $B : (x = L + g(\Delta t)^2/2, t = \Delta t)$. The proper time shown by these clocks is the same

$$\int_0^{\Delta t} dt\sqrt{1 - g^2t^2/c^2} = \Delta t + O(\Delta t^3)$$

Events A and B are simultaneous in frame S but not simultaneous in the accelerated frame S_1 which (at time $t = \Delta t$) has gained a velocity $g\Delta t$ with respect to S.

Let S' be a third frame which moves with constant velocity $v = g\Delta t$ with respect to S. Einstein chooses this frame just so that it is "co-moving" with the accelerated frame S_1 at exactly the S time $t = \Delta t$.

It is simultaneity in S' (which is comoving with S_1), that should be used to decide which two events are simultaneous in the accelerated frame S_1 at this instant. The $t' =$ constant line (which determines the simultaneity in S') passes through event A on the trajectory of C_1 and cuts the trajectory of the second clock C_2 a little further at C. The time shown by clock C_2 kept at L in the accelerated frame S_1 will be the proper time *upto event C while at the same instant (according to S') the time shown by the clock C_1 will be the proper time upto A.*

It is clear from the diagram that the clock C_2 (at event C) will show more time that the clock C_1 at A. We calculate the proper times $\tau(O_1A)$ and $\tau(O_2C)$. $\tau(O_1A)$ is equal upto lowest order to Δt. To find proper time upto C, we need the time coordinate t_C of event C in frame S. The coordinates of these events are as follows in the two frames :

	S	S'
A	$x_A = g(\Delta t)^2/2,\ t = \Delta t$	$t' = t'_A$
C	$x_C = g t_C^2/2,\ t = t_C$	$t' = t'_C = t'_A$

where we have not written those coordinates we do not need.

The basic idea is to calculate t'_A from the first line and then relate $t'_C = t'_A$ to (x_C, t_C). Use Lorentz transformation between S and S' with relative velocity $v = g\Delta t$:

$$\begin{aligned} t'_A &= \gamma(v)\left[\Delta t - \frac{v}{c^2}\frac{g(\Delta t)^2}{2}\right], \qquad (v = g\Delta t) \\ t'_C &= \gamma(v)\left[t_C - \frac{v}{c^2}\left(L + \frac{g(t_C)^2}{2}\right)\right] \end{aligned}$$

Neglecting order $(\Delta t)^3$ terms the first equation gives $t'_A \approx \Delta t$ and so the second gives

$$t'_C = t'_A = \Delta t = t_C - \frac{g\Delta t}{c^2}\left[L + \frac{g(t_C)^2}{2}\right]$$

Solving this equation for t_C and keeping to lowest order we get

$$t_C = \Delta t(1 + gL/c^2)$$

which is also equal to the proper time $\tau(O_2C)$ in this approximation.

Therefore, in the accelerated frame, which feels a constant gravitational field in $-x$ direction , the clock at 'height' L shows a reading $\Delta t(1 + \Phi/c^2)$ when at the same instant (according to this frame) the clock at the origin shows a reading Δt. Here $\Phi = gL$ is the gravitational potential difference between the locations of the clocks.

Exercise 5 $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ are basis vectors in a three dimensional space V and $\{\alpha^1, \alpha^2, \alpha^3\}$ is the corresponding dual basis in V^*. Choose a new basis $\{\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3\}$ where

$$\mathbf{f}_1 = \mathbf{e}_1 + \mathbf{e}_2, \qquad \mathbf{f}_2 = \mathbf{e}_1 - \mathbf{e}_2, \qquad \mathbf{f}_3 = \mathbf{e}_1 + \mathbf{e}_3$$

Find the basis $\{\beta^1, \beta^2, \beta^3\}$ dual to this new basis.

Answer

$$\beta^1 = (\alpha^1 + \alpha^2 - \alpha^3)/2, \beta^2 = (\alpha^1 - \alpha^2 - \alpha^3)/2, \beta^3 = \alpha^3$$

Exercise 6 Show that r vectors $\gamma^1, \ldots, \gamma^r \in V^*, \dim V^* = n$ are linearly dependent if and only if $\gamma^1 \wedge \ldots \wedge \gamma^r = 0$

Answer If $\gamma^1, \ldots, \gamma^r$ are dependent then one can write one of them, say, γ^k as a linear combination of the others. When γ^k is substituted in $\gamma^1 \wedge \ldots \wedge \gamma^r$ the result is zero because each term will have a repeated factor.

On the other hand if $\gamma^1, \ldots, \gamma^r$ are linearly independent then we can take these as the first r vectors of a basis $\gamma^1, \ldots, \gamma^n$ where $\gamma^{r+1}, \ldots, \gamma^n$ are defined appropriately. Let $\mathbf{g}_1, \ldots, \mathbf{g}_n$ be the dual basis in V. Then by definition of the wedge product

$$(\gamma^1 \wedge \ldots \wedge \gamma^r)(\mathbf{g}_1, \ldots, \mathbf{g}_r) = 1$$

therefore $\gamma^1 \wedge \ldots \wedge \gamma^r \neq 0$.

Exercise 7 Kronecker delta
The (1,1) tensor (see section 5.2.5) $\delta \in T^1_1$ is defined in some basis $\{\mathbf{e}_i\}$ and its dual basis $\{\alpha^i\}$ as

$$\delta = \sum \mathbf{e}_i \otimes \alpha^i$$

Show that this definition is independent of the basis used and the tensor has constant components

$$\delta^i_j = 0 \text{ if } i \neq j, \quad \text{and} \quad = 1 \text{ if } i = j$$

Answer

Change a basis and verify that inverse-transpose rule for change of bases makes the definition independent of bases.

Exercise 8 Every (1,1) tensor $t \in T^1_1(V)$determines a linear mapping $T : V \to V$ by the formula (see section 5.2.5)

$$(T\mathbf{v})(\alpha) = t(\alpha, \mathbf{v})$$

What is the linear mapping corrsponding to the Kronecker delta? And what is the linear mapping corresponding to $t = \mathbf{u} \otimes \beta$ for fixed $\mathbf{u} \in V$ and fixed $\beta \in V^*$.

Answer Identity for Kronecker delta. For $t = \mathbf{u} \otimes \beta$ the map T is such that $T(\mathbf{v}) = \beta(\mathbf{v})\mathbf{u}$

Exercise 9 Show that a metric is non-degenerate if and only if the determinant $g = \det g_{ij}$ is non-zero.

Answer

If the g were zero then the symmetric matrix g_{ij} will have an eigenvector with eigenvalue zero. That is, there would be numbers $\lambda_j, j = 1, \ldots, n$ *not all zero* such that for all $i = 1, \ldots, n$ $\sum_j g_{ij}\lambda_j = 0 = \langle \mathbf{e}_i, \sum_j \lambda_j \mathbf{e}_j \rangle$. As $\{\mathbf{e}_i\}$ is a basis it follows that $\langle \mathbf{v}, \sum_j \lambda_j \mathbf{e}_j \rangle = 0$ for every $\mathbf{v}$. Because the inner product is non-degenerate, it follows that $\sum_j \lambda_j \mathbf{e}_j = \mathbf{0}$, and because ,again, $\mathbf{e}_i$ form a basis $\lambda_j = 0$ for all $j = 1, \ldots, n$. This contradicts that not all λ's are zero.

Exercise 10 Show that the inner product for the dual space V^* as defined in §37 is non-degenerate.

Answer

If $\langle \alpha, \beta \rangle$ were to be zero for every choice of β then by choosing β to be equal to members of the basis $\{\nu_j\}$ dual to $\{\mathbf{n}_i\}$, one by one, we can prove that $\alpha(\mathbf{n}_j) = 0$ for all $j = 1, \ldots, n$. This means that $\alpha(\mathbf{v}) = 0$ for any $\mathbf{v} \in V$. Therefore the inner product as defined is non-degenerate.

Exercise 11 Prove that
(1) two non-null, orthogonal vectors in a metric space are linearly independent
(2) two non-orthogonal null vectors are also linearly independent
(3) in (3+1)dimensional Minkowski space, two null vectors if they are orthogonal, then they are necessarily proportional to each other.