SPECIAL RELATIVITY WITH THE K-CALCULUS

Link to: physicspages home page.

To leave a comment or report an error, please use the auxiliary blog.

Reference: d'Inverno, Ray, *Introducing Einstein's Relativity* (1992), Oxford Uni Press. - Sections 2.7, 2.8 and Problems 2.2, 2.3.

The simplest approach to special relativity that I've found is that presented in d'Inverno's book, so I'll run through the argument here and in the next few posts.

d'Inverno uses the rather cryptic title of 'The k-calculus' to introduce special relativity, which makes the reader (well, me, at any rate) think we're in for some high-powered mathematics. In fact, nothing could be further from the truth.

We start, as usual, with the two postulates of relativity:

(1) All inertial observers are equivalent.
(2) The velocity of light is the same in all inertial systems.

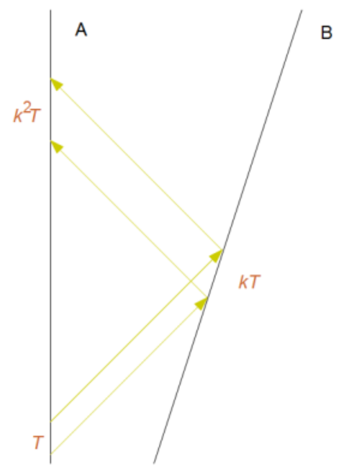
Postulate 1 actually conceals a rather subtle point. Newtonian physics appears to be based on the same principle, since Newton also assumes that the laws of physics must appear the same in all inertial systems (that is, systems moving at constant velocity relative to each other). However, Newton's implicit assumption is that all we need is that if two observers in two different inertial frames perform the same experiments, they should get the same results. Newton does not consider the role of the experimenter in these experiments; that is, that it is necessary to *observe* the experiment in order to interpret its results. This may seem a trivial point, but it is vital, since observation requires receiving signals, usually in the form of light, from the experimental apparatus. Thus the behaviour of light is of crucial importance in interpreting the experiments.

For most experiments, of course, this doesn't make much difference, since we're dealing with speeds much less than the speed of light, so assuming that light has an infinite speed is a good enough approximation. However, as we approach the speed of light, it does in fact make an enormous difference. Newton assumed implicitly that the speed of light would depend on the speed of its source (so that light from a source moving towards you would be faster than light from a stationary source), which contradicts postulate 2 above.

It's usual in relativity to take the speed of light to be 1 ($c=1$), which effectively converts the units of distance into the units of time. Thus, one light-second (the distance light moves in a second) is written simply as 1 second, and so on.

We'll start by considering motion in one spatial dimension only. For such a system, we can represent the state of affairs using a planar space-time diagram, with the x (horizontal) axis being the space dimension and the t (vertical) axis being time. On such a diagram, an observer A who is at rest relative to us would have a constant position, and move only in time. Thus A's world line is vertical, parallel to the t axis.

A second observer B, moving at a constant speed v to the right (relative to A), has a world line at an angle to that of A, as shown in the figure:

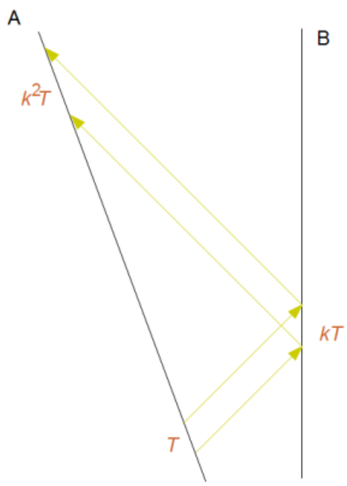


Now suppose A fires off two light pulses, separated by a time T as measured by A. Since $c = 1$, a beam of light's world line always makes an angle of $\pm\pi/4$ with the x axis (positive if the light is moving to the right, negative if to the left). Thus the two pulses emitted by A are as shown by the two yellow lines moving towards the upper right.

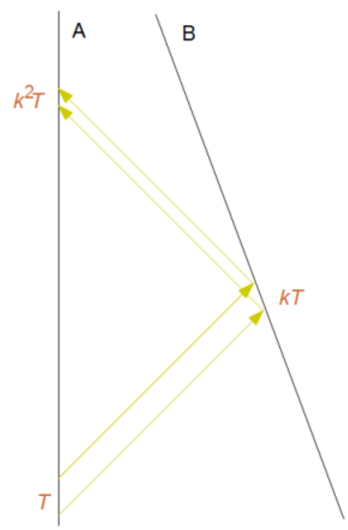
Since B is moving at a constant speed, we make the assumption that the relation between the space and time coordinates of A and B is linear. In that case, we can say that the time interval between the two pulses when they arrive at B is kT (as measured by B), where k is a constant. (This is the 'k' in the k-calculus; I told you it wasn't all that complicated.)

By the same argument, if B reflects these two light pulses back towards A, they will follow the paths shown by the two yellow arrows heading towards the upper left. Because A is moving at a speed of $-v$ relative to B, the same factor should relate the time interval between the reflections at B and their arrival at A, so in terms of the original time interval T, the time interval between the two reflected pulses will be k^2T as measured by A.

We can draw the above diagram from B's point of view:



Finally, we can look at the diagram for the case where B is moving towards A:



Note that in this case, the time intervals get smaller with each interception of the light signals, so it looks like $k < 1$ here.

We can find k in terms of v by making another simple argument. Since we're using light to make our observations, it makes sense to define the distance of an event from an observer as half the time it takes a light signal to make the round trip from the observer to the event and back again. (There wouldn't be much point in defining the distance as the time taken to go from the observer to the event, since the observer has to be able to *see* something to make the measurement, so we need the light signal to get back.) That is, if A sends off a light signal at time t_1 and receives the reflection back at time t_2, then the distance (remember we're using time units for distance) is

$$d = \frac{1}{2}(t_2 - t_1) \tag{1}$$

If we fix the origin of the coordinate system at the observer, then this is also the x coordinate:

$$x = \frac{1}{2}(t_2 - t_1) \tag{2}$$

We can also fix the time of the event to be the average of the emission and reception times, so we get

$$t = \frac{1}{2}(t_1 + t_2) \tag{3}$$

Now suppose that A and B are at the same location at $t = 0$ and synchronize their clocks at that point. Then, after a time T, A sends out a light pulse towards B. As measured by B, this pulse arrives at time kT. (We can think of this as a special case of the two-pulse experiment above, with the first pulse sent at $t = 0$ when both observers are at the same place.) If B reflects this pulse back towards A, it will arrive at A at time k^2T as measured by A. Thus A is able to work out the coordinates of the event where B reflected the pulse, using the above formulas. He gets

$$\begin{aligned} x &= \frac{1}{2}\left(k^2T - T\right) && (4)\\ &= \frac{1}{2}T\left(k^2 - 1\right) && (5)\\ t &= \frac{1}{2}T\left(k^2 + 1\right) && (6) \end{aligned}$$

Since B was at $x = 0$ when $t = 0$ and we now have x and t coordinates for B at a later time, we can work out B's speed:

$$\begin{aligned} v &= \frac{x}{t} && (7)\\ &= \frac{k^2 - 1}{k^2 + 1} && (8) \end{aligned}$$

We can solve this for k, to get

$$\left(k^2+1\right)v = k^2-1 \tag{9}$$

$$k^2 = \frac{1+v}{1-v} \tag{10}$$

$$k = \left(\frac{1+v}{1-v}\right)^{1/2} \tag{11}$$

If $0 < v < 1$ (B is moving to the right away from A), then $k > 1$, while if $-1 < v < 0$ (B moving to the left towards A), then $k < 1$ as we saw in the diagrams above. In fact, it's fairly obvious that if we replace v by $-v$, $k \to 1/k$. In terms of 'red shift' and 'blue shift' Doppler effects, $k > 1$ is a red shift and $k < 1$ a blue shift.

PINGBACKS

Pingback: Composition of velocities in relativity
Pingback: Simultaneity in special relativity

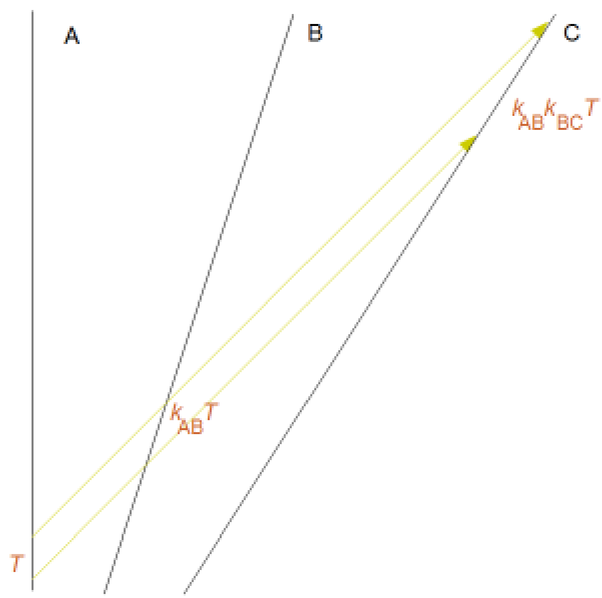
COMPOSITION OF VELOCITIES IN RELATIVITY

Link to: physicspages home page.

To leave a comment or report an error, please use the auxiliary blog.

Reference: d'Inverno, Ray, *Introducing Einstein's Relativity* (1992), Oxford Uni Press. - Section 2.9 and Problems 2.4, 2.5.

The composition of two velocities in special relativity has a particularly simple derivation using the k-calculus. Suppose we have 3 observers as shown in the diagram:



Observer A is at rest relative to us, while B moves to the right with velocity v_{AB} and C also moves to the right with velocity v_{AC}, with both velocities measured relative to A. Now suppose that A emits two light beams separated by a time interval T. From our k-calculus results, we know that B will receive these beams separated by a time $k_{AB}T$. If B then sends these two beams on their way to C, C will receive them at a time interval $k_{BC}(k_{AB}T)$. Thus the overall k-factor from A to C is

$$k_{AC} = k_{AB}k_{BC} \tag{1}$$

We already worked out k in terms of v, so we have

$$k_{AC} = \left(\frac{1+v_{AC}}{1-v_{AC}}\right)^{1/2} \tag{2}$$

$$= \left(\frac{1+v_{AB}}{1-v_{AB}}\right)^{1/2}\left(\frac{1+v_{BC}}{1-v_{BC}}\right)^{1/2} \tag{3}$$

Squaring this equation we get

$$\frac{1+v_{AC}}{1-v_{AC}} = \frac{1+v_{AB}}{1-v_{AB}}\frac{1+v_{BC}}{1-v_{BC}} \tag{4}$$

$$v_{AC}\left(1+\frac{1+v_{AB}}{1-v_{AB}}\frac{1+v_{BC}}{1-v_{BC}}\right) = \frac{1+v_{AB}}{1-v_{AB}}\frac{1+v_{BC}}{1-v_{BC}} - 1 \tag{5}$$

$$v_{AC} = \frac{(1+v_{AB})(1+v_{BC})-(1-v_{AB})(1-v_{BC})}{(1+v_{AB})(1+v_{BC})+(1-v_{AB})(1-v_{BC})} \tag{6}$$

$$= \frac{v_{AB}+v_{BC}}{1+v_{AB}v_{BC}} \tag{7}$$

The composition of two velocities (in the same direction) is therefore less than just the arithmetic sum. In fact, if we start with two velocities, both less than 1 (that is, less than the speed of light), then their sum is also less than 1. We can show this with a little calculus.

Consider the function, defined for $0<x<1$ and $0<y<1$:

$$f(x,y) = \frac{x+y}{1+xy} \tag{8}$$

Taking its two partial derivatives we find

$$\frac{\partial f}{\partial x} = -\frac{y(x+y)}{(1+xy)^2} + \frac{1}{x+y} \tag{9}$$

$$\frac{\partial f}{\partial y} = -\frac{x(x+y)}{(1+xy)^2} + \frac{1}{x+y} \tag{10}$$

Setting each of these to zero, we get the two conditions

$$y^2 = 1 \tag{11}$$

$$x^2 = 1 \tag{12}$$

Thus there are no maxima, minima, or saddle points anywhere inside the region, and the extreme values of the function must lie on the boundary. The boundaries are

$$f(x,1) = \frac{1+x}{1+x} \tag{13}$$

$$= 1 \tag{14}$$

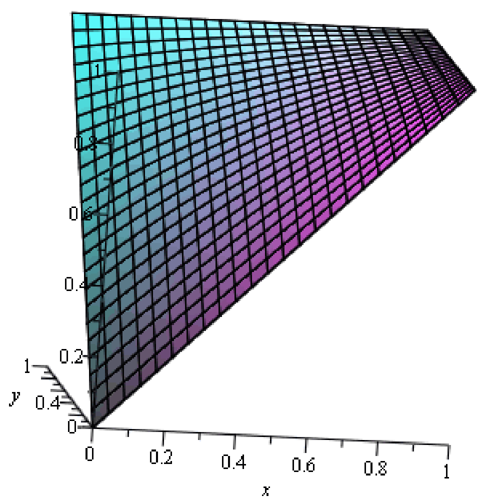
$$f(1,y) = \frac{1+y}{1+y} \tag{15}$$

$$= 1 \tag{16}$$

$$f(x,0) = x \tag{17}$$

$$f(0,y) = y \tag{18}$$

Thus the maximum of the function occurs at the point $(x,y) = (1,1)$ and has the value 1. A plot of the function looks like this:



The nearest corner is the origin, with the point $(1,1)$ lying furthest away.

For velocities $v \ll 1$, the formula reduces to the Newtonian formula. We can approximate the formula above using a Taylor series:

$$\frac{x+y}{1+xy} \simeq (x+y)(1-xy+\ldots) \tag{19}$$

If we save only up to first-order terms, we get

$$\frac{x+y}{1+xy} \simeq x+y \tag{20}$$

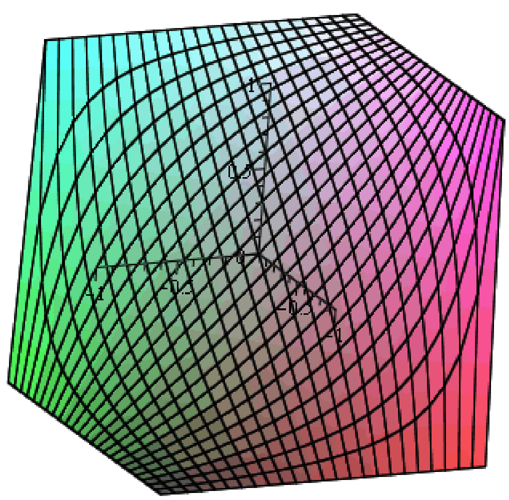
or, in terms of velocities

$$v_{AC} \simeq v_{AB} + v_{BC} \tag{21}$$

For negative velocities, we can look at the region $-1 < x < 1$ and $-1 < y < 1$. The other two boundaries are

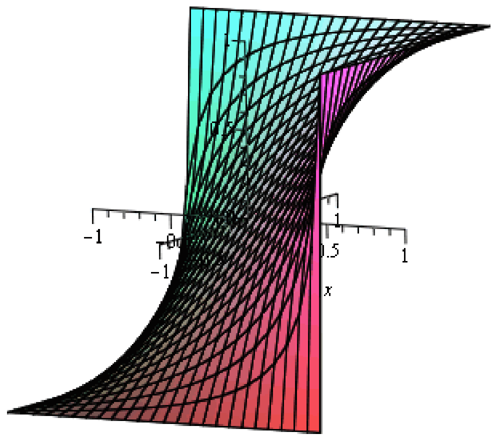
$$
\begin{aligned}
f(x,-1) &= \frac{x-1}{1-x} && (22)\\
&= -1 && (23)\\
f(-1,y) &= \frac{y-1}{1-y} && (24)\\
&= -1 && (25)
\end{aligned}
$$

Thus along these boundaries, the extreme value of the function is -1. The function is actually discontinuous at the two points $(-1,1)$ and $(1,-1)$, where the value tends to ± 1 depending on how you approach the point. A plot looks like this:



The viewpoint is roughly the same as in the previous plot, with the nearest corner being $(-1,-1)$ and the farthest corner at the top being $(1,1)$.

Another plot rotated about $90°$ to the left shows the shape a bit more clearly:


−1
−0.5
0.5
1
−1
1
0.5
x

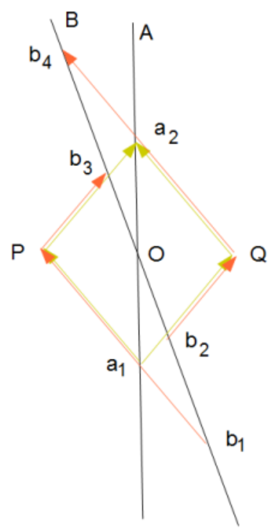
SIMULTANEITY IN SPECIAL RELATIVITY

Link to: physicspages home page.

To leave a comment or report an error, please use the auxiliary blog.

Reference: d'Inverno, Ray, *Introducing Einstein's Relativity* (1992), Oxford Uni Press. - Sections 2.10 and Problem 2.6.

One of the less intuitive predictions of special relativity is that two events which appear to be simultaneous to one observer will not necessarily appear simultaneous to another inertial observer moving relative to the first one. One of the simpler ways of seeing how this comes about is by examining a space-time diagram as shown:



The diagram is shown from the point of view of observer A, who sees the second observer B moving with a speed v to the left. The two observers meet at event O, at which point their clocks are synchronized.

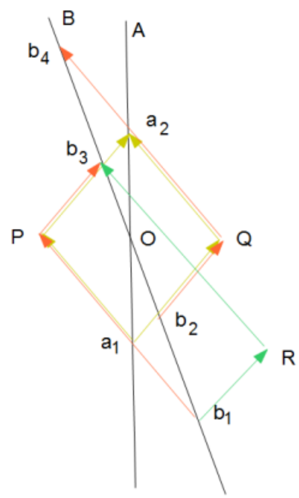
Both observers view two events, P and Q, by reflecting light beams off them. These two events are seen as simultaneous to A, since he emits two light beams at the same time a_1 (yellow arrows). These light beams bounce off P and Q and arrive back at A at the same time a_2. Thus, using our measures of time and distance from the last post, A says that both P and Q are at a distance $\frac{1}{2}(a_2 - a_1)$ and both events occurred at time $\frac{1}{2}(a_2 + a_1)$.

For B to detect event P, he has to send out a light beam at time (in B's frame) b_1 (orange arrow). This beam is reflected by P and arrives back at B at time b_3, so B says the event occurred at $(x_P, t_P) = \left(-\frac{1}{2}(b_3 - b_1), \frac{1}{2}(b_3 + b_1)\right)$.

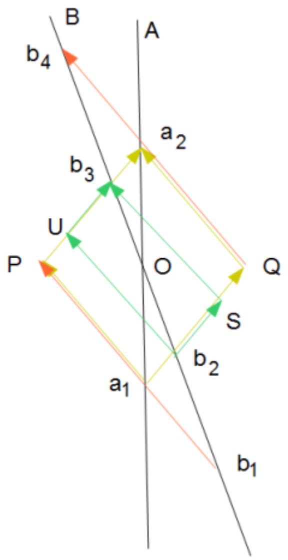
(The minus sign for x_P arises because the light beam is emitted to the left.) Similarly, event Q is detected by emitting a light beam at time b_2 and detecting its reflection at time b_4, so its coordinates appear to B as $(x_Q, t_Q) = \left(\frac{1}{2}(b_4 - b_2), \frac{1}{2}(b_4 + b_2)\right)$.

From the symmetry of the diagram, the time intervals $b_3 - b_1$ and $b_4 - b_2$ are the same, so A and B would agree that the two events P and Q occurred at the same distance on either side of O. However, it's quite clear from the diagram that the midpoint of the time interval b_1 to b_3 (that is, B's measurement of the time of P) is located *before* O and the midpoint of the time interval b_2 to b_4 (that is, B's measurement of the time of Q) is located *after* O, so in B's frame, event P happens first, then A and B meet at event O, then event Q happens.

If we wanted an event R that appeared to B to be simultaneous with event P, we would need an event that could be detected by firing off a light beam at time b_1 and receiving a reflection back at time b_3. Such an event R is shown in this diagram, where the new light beams are drawn in green:

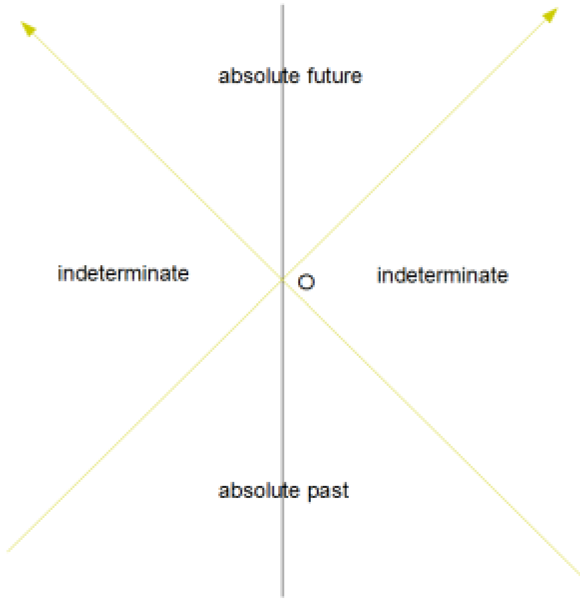


If we wanted an event S that appeared to B to be simultaneous with O, we would need an event that B could detect by firing off a light beam at a time $O - t$, say, and receiving a reflection back at time $O + t$. That is, the sending and receiving of the light beam must occur at equal times on either side of O. Since the times b_2 and b_3 in our diagram above satisfy that condition, we could arrange for event S to be one that receives a light beam from b_2 and reflects it back so that it arrives at B at time b_3 as shown:

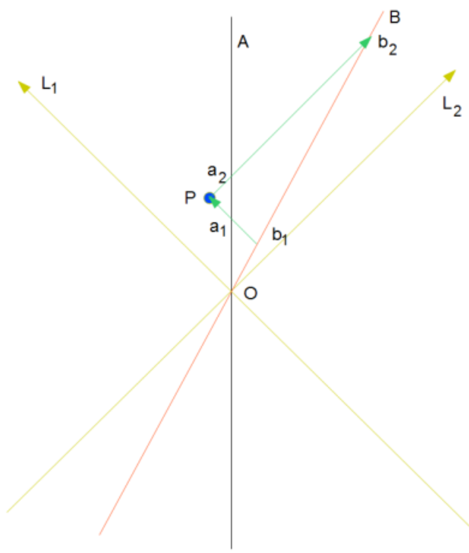


We've added another event U which is also simultaneous with O as seen by B. In fact, any event that lies on the straight line UOS appears to be simultaneous with O as seen by B.

We can use these diagrams to divide up space-time into four regions. If we draw in the light world lines from a given event O, as in the diagram, then any event in the top quadrant is in the absolute future, in the sense that all inertial systems will agree that events in this quadrant occur after O. Similarly, events in the bottom quadrant are in the absolute past, and all observers will agree that they occur before O. Events in the left and right quadrants are indeterminate, in that different observers will disagree about the ordering of these events. Events on the light world lines themselves are also determinate.



To see this, suppose we have an event P (blue dot in following diagram) inside the top quadrant.

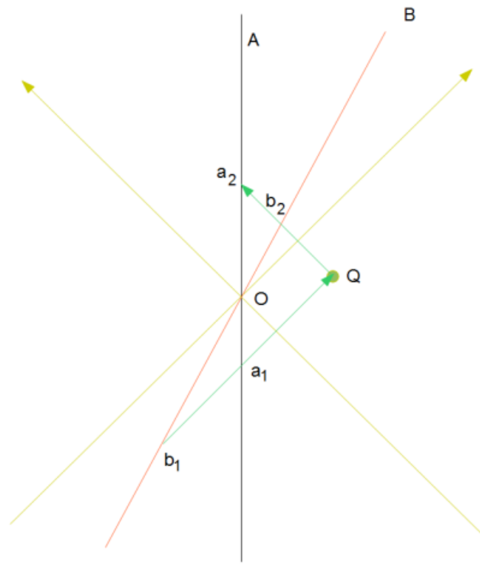


Two observers as shown will observe the event by sending light signals that depart at a_1 and b_1 respectively, and arrive at a_2 and b_2. Because all world lines of observers must lie inside the top quadrant (lines that lie inside the left or right quadrants would represent observers travelling faster than light, which isn't allowed), any light signal from any observer that is to intercept the event has to leave the observer on a world line that is parallel to and above one of the two light lines L_1 and L_2 that define the top quadrant.

Any observer who wishes to detect event O would need to emit signals that lie along lines L_1 or L_2 and receive a response along L_2 or L_1 respectively. These signals always lie below those used to detect event P, so O will always be seen as previous to P.

It's important to note that the two observers will disagree about the actual time at which P occurs, but they will agree that P happens after O.

In the case of an event Q in the right quadrant (see figure below), we can find two observers who disagree about the order in which Q and O occur.



In the diagram, A says that Q occurs at time $\frac{1}{2}(a_1+a_2)$, which is *after* O (since the distance from O to a_1 in the diagram is clearly less than the distance from O to a_2). However, B says that Q occurs at time $\frac{1}{2}(b_1+b_2)$, which clearly occurs before O in B's frame.

If we choose an observer whose world line lies between A and B at just the right angle, we would find that observer saying that O and Q were simultaneous.

MANIFOLDS, CURVES AND SURFACES

Link to: physicspages home page.

To leave a comment or report an error, please use the auxiliary blog.

Reference: d'Inverno, Ray, *Introducing Einstein's Relativity* (1992), Oxford Uni Press. - Sections 5.2-5.3 and Problem 5.1.

Central to general relativity are the mathematical objects known as *tensors*, which are essentially generalizations of vectors. Relativity calculations usually take place on *manifolds*, which are geometric objects that consist of a collection of points, each point satisfying the condition that the area of the manifold near that point resembles Euclidean (that is, 'flat') space. This definition isn't very precise, but for the purposes of relativity, it is probably enough to understand that the general idea is that, although the overall structure of space-time is *not* Euclidean, locally it *seems* Euclidean, which is why Newtonian physics (which assumes Euclidean geometry) works so well on relatively small distance scales and in regions of weak gravitational fields.

The number n of dimensions needed to define a manifold is the number of parameters needed to locate all the points on it. For example, a straight line is a one dimensional manifold, since it can be defined using a single parameter t by the vector equation

$$\mathbf{r}=\mathbf{r}_0+t\mathbf{p}_0 \tag{1}$$

Here, $\mathbf{r}_0=\left(r_{0x},r_{0y}\right)$ is one point on the line, and $\mathbf{p}_0=\left(p_{0x},p_{0y}\right)$ is a vector parallel to the line. The parameter t runs from $-\infty$ to $+\infty$ and in the process, traces out the entire line.

In terms of rectangular coordinates, this equation can be written in parametric form:

$$x = r_{0x}+p_{0x}t \tag{2}$$

$$y = r_{0y}+p_{0y}t \tag{3}$$

We can eliminate t to get a *constraint form* of the equation:

$$t = \frac{1}{p_{0x}}(x - r_{0x}) \tag{4}$$

$$y = r_{0y} + \frac{p_{0y}}{p_{0x}}(x - r_{0x}) \tag{5}$$

$$= \frac{p_{0y}}{p_{0x}}x + \left(r_{0y} - \frac{p_{0y}}{p_{0x}}r_{0x}\right) \tag{6}$$

The last equation is the more familiar slope-intercept form of a line: $y = mx + b$. In this case, the entire manifold is Euclidean, since a straight line is a 'flat' one-dimensional space.

A circle is also a one-dimensional manifold, since again, it needs only a single parameter to trace out the curve. We can write a circle of radius a centred at the origin in parametric form:

$$x = a\cos\theta \tag{7}$$

$$y = a\sin\theta \tag{8}$$

where the parameter θ runs from 0 to 2π. The parameter can be eliminated by squaring each equation and adding to get the constraint form:

$$x^2 + y^2 - a^2 = 0 \tag{9}$$

In this case, the circle itself is non-Euclidean, but if we look at a small enough part of it, it 'resembles' a patch of a one-dimensional Euclidean space (that is, a straight line segment).

A sphere of radius a centred at the origin is a two-dimensional manifold, since it requires two parameters to locate a point on the sphere, and if we restrict ourselves to a small enough patch on the sphere, it resembles a portion of a plane, which is a two-dimensional Euclidean space. The parametric equations of the sphere are:

$$x = a\sin\theta\cos\phi \tag{10}$$

$$y = a\sin\theta\sin\phi \tag{11}$$

$$z = a\cos\theta \tag{12}$$

Here, θ is the usual angle between the z axis and the radius vector, and ϕ is the angle between the x axis and the projection of the radius vector into the xy plane. These two parameters can be eliminated by squaring and adding all three equations to get the constraint

$$x^2 + y^2 + z^2 - a^2 = 0 \tag{13}$$

More generally, we can consider a manifold of any number n of dimensions and look at curves and surfaces within that manifold. For an n-dimensional manifold, we need n coordinates to specify any given point. For reasons that will become apparent later (that is, in a future post), tensor theory writes these coordinates with a superscript, so they are denoted $x^1, x^2 \ldots x^{n-1} x^n$. Obviously, there is the potential of confusing these superscripts with exponents, so we need to be careful with the notation. If we want to write an exponent, it's usual to enclose the coordinate in parentheses, so that $\left(x^1\right)^2$ is the square of coordinate x^1.

To specify a curve or surface within the manifold, we need to know the dimension of the curve or surface. (Well, a curve is always defined as one-dimensional, but the term 'surface' can have any number of dimensions from 2 up to n.) For a subsurface (where the number of dimensions m is strictly less than n) we need m parameters u^i to define it, so we get

$$x^a = x^a\left(u^1, u^2, \ldots, u^m\right) \tag{14}$$

where this represents n equations, one for each $a \in \{1, 2, \ldots, n\}$. This is the generalization of the examples above, in which we embedded a one-dimensional subsurface (the circle) in a two-dimensional manifold (a plane), and a two-dimensional subsurface (the sphere) in a three-dimensional manifold (Euclidean 3-d space).

If we have m dimensions in the subsurface embedded in an n-dimensional manifold, we must have $n - m$ constraints which can be written in the form

$$f^b\left(x^1, x^2, \ldots, x^n\right) = 0; \quad b = 1 \ldots m - n \tag{15}$$

We saw the constraint equations in the examples above. These constraint equations can be obtained by eliminating the parameters u^i from the parametric equations, as we did above.

PINGBACKS

Pingback: Event horizon: distance from external radius up to r = 2GM
Pingback: Coordinate transformations - the Jacobian determinant
Pingback: Contravariant tensors

COORDINATE TRANSFORMATIONS - THE JACOBIAN DETERMINANT

Link to: physicspages home page.

To leave a comment or report an error, please use the auxiliary blog.

Reference: d'Inverno, Ray, *Introducing Einstein's Relativity* (1992), Oxford Uni Press. - Section 5.4 and Problem 5.2.

Since one of the main aspects of the definition of a tensor is the way it transforms under a change in coordinate systems, it's important to consider how such coordinate changes work.

We'll consider two coordinate systems, one denoted by unprimed symbols x^i and the other by primed symbols x'^i. In general, one system is a function of the other one, so we can write

$$x'^i = x'^i(x) \tag{1}$$

where the index i runs over the n dimensions of the manifold (so we have a set of n equations), and the symbol x without an index means the set of all components of x, so it's equivalent to (but shorter than) writing

$$x'^i = x'^i\left(x^1, x^2, x^3, \ldots, x^n\right) \tag{2}$$

Now suppose we want to do an integral over a portion of the manifold that is bounded by some subsurface in the manifold. As we know from elementary calculus, the differential volume (or area) has a different form depending on which coordinate system we're using. For example, in 3-d rectangular coordinates, the volume element is $dxdydz$, while in spherical coordinates it is $r^2 \sin\theta dr d\theta d\phi$.

To see how this works we can start with one dimension. If we have an integral in rectangular coordinates such as

$$\int_{x_1}^{x_2} f(x)dx \tag{3}$$

we can change coordinate systems if we define $x = x(u)$. Then we have $dx = \frac{dx}{du}du$. To transform the limits of the integral, we need to invert the definition to get $u = u(x)$. Then the integral becomes

$$\int_{u(x_1)}^{u(x_2)} f(x(u)) \frac{dx}{du} du \tag{4}$$

Essentially, this redefines the line element into the u coordinate system.

In two dimensions, we'd start off with (we'll leave out the limits on the integrals since we're really interested only in the area element):

$$\int\int f(x,y)dxdy \tag{5}$$

Now if we want to switch to another coordinate system, we define

$$u = u(x,y) \tag{6}$$

$$v = v(x,y) \tag{7}$$

Consider now an elemental rectangle R in the xy plane. The rectangle has its lower left corner at the point (x_0, y_0) and has dimensions Δx and Δy, so that its area is $\Delta x \Delta y$.

We want to see how this rectangle transforms under the coordinate transformation above. The new elemental area will not necessarily be a rectangle, but we can transform it point by point to get the new shape. Starting with the lower left corner, this transforms to

$$(u_0, v_0) = [u(x_0, y_0), v(x_0, y_0)] \tag{8}$$

We can write the general transformation as a vector:

$$\mathbf{r}(x,y) = u(x,y)\hat{\mathbf{i}} + v(x,y)\hat{\mathbf{j}} \tag{9}$$

Here, $\mathbf{r}$ is the transformed location of the original point (x,y), written with respect to the rectangular basis vectors.

The idea now is to consider what happens as Δx and Δy tend to zero. In this case, the transformed version of R tends to a parallelogram whose sides are parallel to the transformation of the two sides of R that touch at the point $P_0 = (x_0, y_0)$ (the lower left corner of R we mentioned above). Consider first the edge of R along the line $y = y_0$ (the bottom of the rectangle). We can think of this edge as a tangent to the rectangle at the point P_0. How does this tangent transform?

Well, the lower edge of R transforms as

$$\mathbf{r}(x, y_0) = u(x, y_0)\hat{\mathbf{i}} + v(x, y_0)\hat{\mathbf{j}} \tag{10}$$

The tangent along this curve is then the derivative with respect to x, so we get

$$\frac{\partial}{\partial x}\mathbf{r}\left(x,y_0\right)=\frac{\partial}{\partial x}u\left(x,y_0\right)\hat{\mathbf{i}}+\frac{\partial}{\partial x}v\left(x,y_0\right)\hat{\mathbf{j}} \tag{11}$$

Thus the tangent along the bottom edge of R at the transformed location of P_0 is

$$\mathbf{r}_x\equiv\frac{\partial}{\partial x}\mathbf{r}\left(x,y_0\right)\bigg|_{x=x_0}=\frac{\partial}{\partial x}u\left(x,y_0\right)\bigg|_{x=x_0}\hat{\mathbf{i}}+\frac{\partial}{\partial x}v\left(x,y_0\right)\bigg|_{x=x_0}\hat{\mathbf{j}} \tag{12}$$

By the same argument, the tangent at P_0 along the left edge of R is found by setting $x=x_0$ and differentiating with respect to y, and we get

$$\mathbf{r}_y\equiv\frac{\partial}{\partial y}\mathbf{r}\left(x_0,y\right)\bigg|_{y=y_0}=\frac{\partial}{\partial y}u\left(x_0,y\right)\bigg|_{y=y_0}\hat{\mathbf{i}}+\frac{\partial}{\partial y}v\left(x_0,y\right)\bigg|_{y=y_0}\hat{\mathbf{j}} \tag{13}$$

By the definition of a derivative, we can write these tangents in the form

$$\mathbf{r}_x = \lim_{\Delta x\to 0}\frac{\mathbf{r}\left(x_0+\Delta x,y_0\right)-\mathbf{r}\left(x_0,y_0\right)}{\Delta x} \tag{14}$$

$$\mathbf{r}_y = \lim_{\Delta y\to 0}\frac{\mathbf{r}\left(x_0,y_0+\Delta y\right)-\mathbf{r}\left(x_0,y_0\right)}{\Delta y} \tag{15}$$

The vector $\mathbf{r}\left(x_0+\Delta x,y_0\right)-\mathbf{r}\left(x_0,y_0\right)$ connects the transformed lower left corner of R to the transformed lower right corner. Similarly $\mathbf{r}\left(x_0,y_0+\Delta y\right)-\mathbf{r}\left(x_0,y_0\right)$ connects the lower left corner to the upper left corner. Thus these two vectors define the sides of a parallelogram that, for very small Δx and Δy, is a good approximation to the transformed R. In this approximation, we can write

$$\mathbf{r}\left(x_0+\Delta x,y_0\right)-\mathbf{r}\left(x_0,y_0\right) \simeq \mathbf{r}_x\Delta x \tag{16}$$

$$\mathbf{r}\left(x_0,y_0+\Delta y\right)-\mathbf{r}\left(x_0,y_0\right) \simeq \mathbf{r}_y\Delta y \tag{17}$$

The area of a parallelogram is $A=s_1s_2\sin\theta$, where s_1 and s_2 are two adjacent sides and θ is the angle between them. If we have two vectors corresponding to the sides, the area is thus the magnitude of the cross product of the vectors. So we get

$$\Delta A=\left|\mathbf{r}_x\times\mathbf{r}_y\right|\Delta x\Delta y \tag{18}$$

Using the equations above, we can work out this cross product. We'll use the notation $u_y\equiv\partial u/\partial y$ to save space. We get

$$\mathbf{r}_x \times \mathbf{r}_y = \begin{vmatrix} \hat{\mathbf{i}} & \hat{\mathbf{j}} & \hat{\mathbf{k}} \\ u_x & v_x & 0 \\ u_y & v_y & 0 \end{vmatrix} = (u_x v_y - v_x u_y)\,\hat{\mathbf{k}} \tag{19}$$

The coefficient of $\hat{\mathbf{k}}$ is itself a 2×2 determinant, and can be written as

$$J(x,y) \equiv \begin{vmatrix} u_x & u_y \\ v_x & v_y \end{vmatrix} \tag{20}$$

This is called the *Jacobian* of the transformation. The area element is thus

$$dA = J\,(x,y)\,dxdy \tag{21}$$

Now this is all very well, but the differentials Δx and Δy are still in the original coordinate system. How can we use this result to transform the integral that we began with?

The trick is to assume that the transformation is invertible, that is, that we can also write

$$x = x\,(u,v) \tag{22}$$

$$y = y\,(u,v) \tag{23}$$

We can run through the same argument again to get

$$dA = J\,(u,v)\,dudv \tag{24}$$

with

$$J\,(u,v) = \begin{vmatrix} x_u & x_v \\ y_u & y_v \end{vmatrix} \tag{25}$$

That is:

$$\int\int f(x,y)dxdy = \int\int f\left[x\,(u,v)\,,y\,(u,v)\right]\left|J\,(u,v)\right|dudv \tag{26}$$

Note that we've taken the absolute value of J since we're dealing with an area element, which must be positive.

It can also be shown that (the proof would make this post too long) the Jacobian satisfies a very convenient property:

$$J(u,v) = \frac{1}{J(x,y)} \tag{27}$$

That is, the Jacobian of an inverse transformation is the reciprocal of the Jacobian of the original transformation.

The Jacobian generalizes to any number of dimensions, so we get, reverting to our primed and unprimed coordinates:

$$J\left(x'\right) = \begin{vmatrix} \frac{\partial x^1}{\partial x'^1} & \frac{\partial x^1}{\partial x'^2} & \cdots & \frac{\partial x^1}{\partial x'^n} \\ \frac{\partial x^2}{\partial x'^1} & \frac{\partial x^2}{\partial x'^2} & \cdots & \frac{\partial x^2}{\partial x'^n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x^n}{\partial x'^1} & \frac{\partial x^n}{\partial x'^2} & \cdots & \frac{\partial x^n}{\partial x'^n} \end{vmatrix} \tag{28}$$

For obvious reasons, this can be abbreviated to

$$J = \left| \frac{\partial x^a}{\partial x'^b} \right| \tag{29}$$

As a simple example, consider the transformation from rectangular to polar coordinates in 2-d. From the above, the Jacobian we want is $J(r,\theta)$ which requires expressing the old coordinates in terms of the new ones. The transformation is

$$x = r\cos\theta \tag{30}$$

$$y = r\sin\theta \tag{31}$$

So we have

$$J(r,\theta) = \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix} = r \tag{32}$$

Thus the transformation of the area element is

$$dxdy \rightarrow rdrd\theta \tag{33}$$

For the inverse transformation, we have

$$r = \sqrt{x^2+y^2} \tag{34}$$

$$\theta = \tan^{-1}\frac{y}{x} \tag{35}$$

so

$$J(x,y)=\begin{vmatrix}\frac{x}{\sqrt{x^2+y^2}} & \frac{y}{\sqrt{x^2+y^2}}\\ \frac{-y/x^2}{1+(y/x)^2} & \frac{1/x}{1+(y/x)^2}\end{vmatrix}=\frac{1}{\sqrt{x^2+y^2}}=\frac{1}{r}\tag{36}$$

Thus $J(u,v)=1/J(x,y)$ as required.

In 3-d,

$$x = r\sin\theta\cos\phi\tag{37}$$
$$y = r\sin\theta\sin\phi\tag{38}$$
$$z = r\cos\theta\tag{39}$$

$$J(r,\theta,\phi)=\begin{vmatrix}\sin\theta\cos\phi & r\cos\theta\cos\phi & -r\sin\theta\sin\phi\\ \sin\theta\sin\phi & r\cos\theta\sin\phi & r\sin\theta\cos\phi\\ \cos\theta & -r\sin\theta & 0\end{vmatrix}=r^2\sin\theta\tag{40}$$

For the inverse:

$$r = \sqrt{x^2+y^2+z^2}\tag{41}$$
$$\theta = \tan^{-1}\frac{\sqrt{x^2+y^2}}{z}\tag{42}$$
$$\phi = \tan^{-1}\frac{y}{x}\tag{43}$$

$$J(x,y,z)=\begin{vmatrix}\frac{x}{\sqrt{x^2+y^2+z^2}} & \frac{y}{\sqrt{x^2+y^2+z^2}} & \frac{z}{\sqrt{x^2+y^2+z^2}}\\ \frac{x/\left(z\sqrt{x^2+y^2}\right)}{1+\left(x^2+y^2\right)/z^2} & \frac{y/\left(z\sqrt{x^2+y^2}\right)}{1+\left(x^2+y^2\right)/z^2} & \frac{-\sqrt{x^2+y^2}/z^2}{1+\left(x^2+y^2\right)/z^2}\\ \frac{-y/x^2}{1+(y/x)^2} & \frac{1/x}{1+(y/x)^2} & 0\end{vmatrix}\tag{44}$$

Converting back to spherical coordinates proves a bit easier. Substituting the above transformation equations, along with

$$r^2\sin^2\theta=x^2+y^2\tag{45}$$

helps to simplify things.

$$J(x,y,z)=\begin{vmatrix}\frac{x}{r} & \frac{y}{r} & \frac{z}{r}\\ \frac{xz}{r^3\sin\theta} & \frac{yz}{r^3\sin\theta} & \frac{-\sin\theta}{r}\\ \frac{-y}{r^2\sin^2\theta} & \frac{x}{r^2\sin^2\theta} & 0\end{vmatrix}\tag{46}$$

The determinant now comes out to

(47)
$$J(x,y,z) = \frac{-y}{r^2\sin^2\theta}\left(\frac{-\sin\theta}{r}\frac{y}{r}-\frac{z}{r}\frac{yz}{r^3\sin\theta}\right)-\frac{x}{r^2\sin^2\theta}\left(\frac{-\sin\theta}{r}\frac{x}{r}-\frac{z}{r}\frac{xz}{r^3\sin\theta}\right)$$
(48)
$$= \frac{1}{r^4\sin^2\theta}\left(y^2\left(\sin\theta+\frac{z^2}{r^2\sin\theta}\right)+\left(x^2\left(\sin\theta+\frac{z^2}{r^2\sin\theta}\right)\right)\right)$$
(49)
$$= \frac{x^2+y^2}{r^4\sin^2\theta}\left(\sin\theta+\frac{z^2}{r^2\sin\theta}\right)$$
(50)
$$= \frac{1}{r^2}\left(\sin\theta+\frac{z^2}{r^2\sin\theta}\right)$$
(51)
$$= \frac{1}{r^2}\left(\frac{r^2\sin^2\theta+z^2}{r^2\sin\theta}\right)$$
(52)
$$= \frac{1}{r^2}\left(\frac{x^2+y^2+z^2}{r^2\sin\theta}\right)$$
(53)
$$= \frac{1}{r^2\sin\theta}$$

PINGBACKS

Pingback: Contravariant tensors
Pingback: Covariant and mixed tensors
Pingback: Noether's theorem and conservation laws

SUMMATION CONVENTION

Link to: physicspages home page.

To leave a comment or report an error, please use the auxiliary blog.

Reference: d'Inverno, Ray, *Introducing Einstein's Relativity* (1992), Oxford Uni Press. - Section 5.4 and Problems 5.3, 5.4.

We've met the summation convention briefly before, but we'll examine it again, since it is central to relativity calculations.

Many formulas involving coordinate systems require the summation over the various coordinates. For example, suppose we have a function of n independent variables x^i (where $i = 1, \cdots, n$):

$$f = f(x) \tag{1}$$

If this function defines an m-dimensional subsurface within the n-dimensional manifold, we need m parameters to describe the subsurface. That means that we can write the function in parametric form, where each of the x^i is a function of m parameters u^j (where $j = 1, \cdots, m$). That is

$$f = f(x(u)) \tag{2}$$

The derivative of this function with respect to one of the parameters can be found using the chain rule:

$$\frac{\partial f}{\partial u^a} = \sum_{i=1}^{n} \frac{\partial f}{\partial x^i} \frac{\partial x^i}{\partial u^a} \tag{3}$$

The index i is repeated in the summand, with one occurrence being 'on the top' and one 'on the bottom'. Einstein noticed that this pattern occurs regularly in relativity theory: whenever an index is repeated with one occurrence up and the other down, this index is summed over. As such, the summation sign is redundant and, since it's a pain to have to write it over and over, it can be dropped without any ambiguity. Thus the above formula can be written

$$\frac{\partial f}{\partial u^a} = \frac{\partial f}{\partial x^i} \frac{\partial x^i}{\partial u^a} \tag{4}$$

As we'll see when we study tensors properly, a tensor can have upper and lower indices (the upper ones are *contravariant* and the lower ones are *covariant*), and terms involving products of tensors use the same summation convention. Thus we have, as examples

$$A_b^a B^{bc} = \sum_{b=1}^{n} A_b^a B^{bc} = T^{ac} \tag{5}$$

$$C_l^{ijk} D_i^p E_j^{qr} = \sum_{i=1}^{n}\sum_{j=1}^{n} C_l^{ijk} D_i^p E_j^{qr} = S_l^{kpqr} \tag{6}$$

The last terms in these examples show that we can write the result of these implied sums as a single tensor that contains only the non-repeated indices, so that T^{ac} and S_l^{kpqr} are the results of the sums.

A repeated index is known as a *dummy index*, since it is just a summation label, and it can be changed to any symbol (that doesn't appear as another index) without affecting the formula. So we could write

$$A_b^a B^{bc} = A_i^a B^{ic} = A_t^a B^{tc} = \cdots \tag{7}$$

Note that we *cannot* replace a non-repeated index, since it doesn't get summed over and appears in the final result. Thus we couldn't replace a or c in the last example.

As an example of the interchangeability of indices, consider the expression

$$(Z_{abc} + Z_{cab} + Z_{bca}) X^a X^b X^c \tag{8}$$

Since all three indices are repeated, they are all dummies and can be changed. Further, the factor $X^a X^b X^c$ is symmetric with respect to the three indices, so it doesn't matter what order the Xs are written. Multiplying the terms out we get

$$(Z_{abc} + Z_{cab} + Z_{bca}) X^a X^b X^c = Z_{abc} X^a X^b X^c + Z_{cab} X^a X^b X^c + Z_{bca} X^a X^b X^c \tag{9}$$

In the second term we can relabel the indices as $a \to b$, $b \to c$ and $c \to a$, and in the third term we can relabel as $a \to c$, $b \to a$ and $c \to b$. As a result, all three terms are equal, and we get

$$(Z_{abc} + Z_{cab} + Z_{bca}) X^a X^b X^c = 3 Z_{abc} X^a X^b X^c \tag{10}$$

We've used the Kronecker delta many times in these posts. It's defined as

$$\delta_b^a = \begin{cases} 1 & a = b \\ 0 & a \neq b \end{cases} \tag{11}$$

Many uses of the Kronecker delta are sloppy with the indices, in that they usually both appear at the bottom as in δ_{ab}. In relativity, it's important to write the indices in the correct location so that if the symbol appears in a summation, the correct summation is performed. For example

$$\delta_a^b X^a = \sum_{a=1}^{n} \delta_a^b X^a = X^b \tag{12}$$

$$\delta_a^b X_b = \sum_{b=1}^{n} \delta_a^b X_b = X_a \tag{13}$$

If an summation expression contains several deltas, only those terms where all the deltas are non-zero survive. Thus

$$\delta_a^b \delta_b^c \delta_c^d = \delta_a^d \tag{14}$$

This follows, since the indices b and c are summed, and in the first delta the only surviving term is when $a = b$, and in the last delta, we must have $c = d$. The middle delta requires $b = c$, so combining this with the other two results gives us $a = d$ as the only surviving term. Again, note that it's only the non-repeated indices that survive the summation process.

PINGBACKS

Pingback: Contravariant tensors

CONTRAVARIANT TENSORS

Link to: physicspages home page.

To leave a comment or report an error, please use the auxiliary blog.

Reference: d'Inverno, Ray, *Introducing Einstein's Relativity* (1992), Oxford Uni Press. - Section 5.5 and Problems 5.5, 5.6.

Suppose we have a function defined in an n-dimensional manifold:

$$f=f(x) \tag{1}$$

where x represents all n coordinates in some coordinate system.

Now suppose we define some curve within the manifold, and ask how the function varies as we move along this curve. Since the curve is one-dimensional, it can be described using a single parameter u. Then the rate of change of f along u can be found using the chain rule:

$$\frac{df}{du}=\frac{\partial f}{\partial x^i}\frac{dx^i}{du} \tag{2}$$

where we're using the summation convention. We've used the total derivative notation for the derivatives with respect to u, since in both these cases, we're considering a derivative along the particular path that is given by the single parameter u, so there is only one independent variable in those cases. The derivative with respect to x^i must be partial, since f depends on (in general) all the x^is.

In particular, we can consider a change of coordinates, and write each of the new, primed coordinates as a function of the old unprimed coordinates, like so:

$$x'^a=x'^a(x) \tag{3}$$

Using the chain rule formula, we find that

$$\frac{dx'^a}{du}=\frac{\partial x'^a}{\partial x^i}\frac{dx^i}{du} \tag{4}$$

Dropping the du off both sides (the way physicists do, and mathematicians hate), we get an expression for the transformation of the differentials between two coordinate systems:

$$dx'^a = \frac{\partial x'^a}{\partial x^i} dx^i \tag{5}$$

Any quantity that transforms in this way is called a *contravariant tensor of rank 1*, or, for short, a *contravariant vector*.

The tangent vector $\mathbf{t}(u)$ (from elementary calculus) to a parametric curve given in vector form is the derivative of each component of the curve's vector with respect to u, and has components in a given coordinate system:

$$t^a(u) = \frac{dx^i}{du} \tag{6}$$

Thus the tangent vector is a contravariant vector.

As an example, suppose we have a circle of radius a centred at the origin. We can write this in rectangular coordinates as a curve using the angle θ as the parameter in the usual way:

$$x = a\cos\theta \tag{7}$$
$$y = a\sin\theta \tag{8}$$

In this system, the tangent vector is then

$$t(\theta) = \frac{d}{d\theta}\begin{pmatrix} x & y \end{pmatrix} = \begin{pmatrix} -a\sin\theta & a\cos\theta \end{pmatrix} = \begin{pmatrix} -y & x \end{pmatrix} \tag{9}$$

This makes sense, since the tangent to a circle is perpendicular to the radius vector, and the radius vector has coordinates $\begin{pmatrix} x & y \end{pmatrix}$, so the dot product of radius and tangent gives zero as required.

In polar coordinates (which we'll take as the primed system), the tangent is exceptionally easy to find, since the parametric equations are

$$r = a \tag{10}$$
$$\theta = \theta \tag{11}$$

so the tangent is

$$t'(\theta) = \frac{d}{d\theta}\begin{pmatrix} r & \theta \end{pmatrix} = \begin{pmatrix} 0 & 1 \end{pmatrix} \tag{12}$$

However, we can use the tensor transformation equation above to check this. We've already done the calculation of the transformation matrix when we considered Jacobians earlier, so we get

$$r = \sqrt{x^2+y^2} \tag{13}$$

$$\theta = \tan^{-1}\frac{y}{x} \tag{14}$$

so

$$\frac{\partial x'^a}{\partial x^i} = \begin{bmatrix} \frac{x}{\sqrt{x^2+y^2}} & \frac{y}{\sqrt{x^2+y^2}} \\ \frac{-y/x^2}{1+(y/x)^2} & \frac{1/x}{1+(y/x)^2} \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\frac{\sin\theta}{a} & \frac{\cos\theta}{a} \end{bmatrix} \tag{15}$$

If we multiply the terms out, we get

$$t'(\theta) = \begin{bmatrix} \cos\theta & \sin\theta \\ -\frac{\sin\theta}{a} & \frac{\cos\theta}{a} \end{bmatrix} \begin{bmatrix} -a\sin\theta \\ a\cos\theta \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{16}$$

Contravariant tensors of higher rank can be defined by a simple extension of the above formula. A rank-2 tensor transforms according to:

$$X'^{ab} = \frac{\partial x'^a}{\partial x^i}\frac{\partial x'^b}{\partial x^j}X^{ij} \tag{17}$$

Higher ranks are defined by just adding in more transformation matrixes, as you'd expect.

We can define the product of two contravariant vectors as

$$T^{ab} = X^a X^b \tag{18}$$

Note that this is *not* a dot or cross product, since there are no repeated indices on the right, so no implied summation. This equation is shorthand for an $n \times n$ matrix with components given by the equation tiself. For example, the top row of the matrix has elements X^1X^1, $X^1X^2, \ldots X^1X^n$, with similar results for the other rows.

This quantity transforms by applying the formula above for vectors:

$$T'^{ab} = X'^a X'^b \tag{19}$$

$$= \left(\frac{\partial x'^a}{\partial x^i}X^i\right)\left(\frac{\partial x'^b}{\partial x^j}X^j\right) \tag{20}$$

$$= \frac{\partial x'^a}{\partial x^i}\frac{\partial x'^b}{\partial x^j}X^iX^j \tag{21}$$

Thus the product of two vectors has the required transformation property. By the same argument, the product of m contravariant vectors is a rank-m contravariant tensor.

PINGBACKS

Pingback: Inertia tensor
Pingback: Covariant and mixed tensors
Pingback: Kronecker delta as a tensor
Pingback: Higher order derivatives are not tensors
Pingback: Tensor arithmetic
Pingback: Tangent space
Pingback: Notation for Relativistic Quantum Mechanics

COVARIANT AND MIXED TENSORS

Link to: physicspages home page.

To leave a comment or report an error, please use the auxiliary blog.

Reference: d'Inverno, Ray, *Introducing Einstein's Relativity* (1992), Oxford Uni Press. - Section 5.6 and Problem 5.7.

We've seen that objects such as the tangent vector to a curve are contravariant tensors, in that they transform under a change of coordinates according to the rule (for a rank-2 tensor, for example):

$$X'^{ab} = \frac{\partial x'^a}{\partial x^i}\frac{\partial x'^b}{\partial x^j}X^{ij} \tag{1}$$

Note that the indices on the tensor are superscripts, and in the transformation, the original coordinate system's components are those with which the derivative is taken with respect to. That is, the *new* (primed) coordinates are taken to be functions of the *old* (unprimed) coordinates.

Suppose we turn the tables and express the unprimed coordinates as functions of the primed ones, like so:

$$x^a = x^a\left(x'\right) \tag{2}$$

where as usual this notation indicates a set of n equations, one for each of the x^a, and the argument of the function indicates that it is a function of all the primed coordinates. This is the inverse of the original transformation from unprimed to primed coordinates.

Now suppose we have a function defined in terms of the unprimed coordinates:

$$g = g\left(x\right) \tag{3}$$

We can write this as a function of the primed coordinates using the transformation equations above:

$$g = g\left(x\left(x'\right)\right) \tag{4}$$

When we derived the condition for a contravariant tensor, we considered a one-dimensional curve defined within the manifold by using a single parameter u, and then we asked how the function changed as we moved along this curve. This time, we ask simply for the derivative of a function with respect to each of the primed coordinates. Using the chain rule, we get

$$\frac{\partial g}{\partial x'^a} = \frac{\partial g}{\partial x^i}\frac{\partial x^i}{\partial x'^a} \tag{5}$$

That is, the quantity $\partial g/\partial x^i$ transforms by multiplying it with the term $\partial x^i/\partial x'^a$ and summing over i. The quantities $\partial x^i/\partial x'^a$ are entries in the Jacobian determinant for the inverse transformation. Furthermore the index i in $\partial g/\partial x^i$ is now on the bottom (it is a superscript index, but it's in the denominator, so it counts as a lower index). We can write any object that transforms in this way using the notation T_a, so that

$$T'_a = \frac{\partial x^i}{\partial x'^a} T_i \tag{6}$$

This is called a *covariant vector,* or *covariant tensor of rank 1*. Higher rank tensors can be defined in the usual way, by multiplying by further derivative factors. Thus a rank 2 covariant tensor transforms as

$$T'_{ab} = \frac{\partial x^i}{\partial x'^a}\frac{\partial x^j}{\partial x'^b} T_{ij} \tag{7}$$

and so on.

We can also define mixed tensors (tensors that contain both contravariant and covariant indexes) in a relatively obvious way. For example, a tensor with contravariant rank 2 and covariant rank 1, written as a (2,1) tensor, is defined by

$$T'^{ab}{}_{c} = \frac{\partial x'^a}{\partial x^i}\frac{\partial x'^b}{\partial x^j}\frac{\partial x^k}{\partial x'^c} T^{ij}{}_{k} \tag{8}$$

Note the position of the primed and unprimed coordinates in each case. The summation convention applies only to repeated indexes where one index in each pair is upper and the other is lower.

Incidentally, the relative positioning of the indexes in the tensor symbol seems to be largely a matter of taste. That is, some (well, most, actually) books write the indexes so that there is a space in the bottom to avoid overlap with the top indexes (as in $T^{ab}{}_{c}$) while other books leave out the space, so we get T^{ab}_{c}. Since a tensor is defined in terms of its transformation equation, and the order in which we write the derivatives doesn't matter, it shouldn't matter much whether we insert a space in the notation or not, since the order of the indexes doesn't matter. What *is* important is which indexes are upper (contravariant) and which are lower (covariant).

PINGBACKS

Pingback: Gradient as covector: example in 2-d
Pingback: Christoffel symbols
Pingback: Kronecker delta as a tensor
Pingback: Higher order derivatives are not tensors
Pingback: Tensor arithmetic
Pingback: Covariant derivative of covariant vector

KRONECKER DELTA AS A TENSOR

Link to: physicspages home page.

To leave a comment or report an error, please use the auxiliary blog.

Reference: d'Inverno, Ray, *Introducing Einstein's Relativity* (1992), Oxford Uni Press. - Section 5.6 and Problem 5.8.

The Kronecker delta δ_b^a is actually a tensor, as it transforms according to the rules for such a tensor, as we'll see. Recall that the rule for transformation of a contravariant vector(rank-1 tensor) is

$$T'^a = \frac{\partial x'^a}{\partial x^i} T^i \tag{1}$$

and for a covariant vector:

$$T'_a = \frac{\partial x^i}{\partial x'^a} T_i \tag{2}$$

For a mixed tensor, we multiply by the right combination of derivatives to effect a coordinate transformation. If δ_b^a really is a tensor, then it should transform properly. The numerical value of δ_b^a is

$$\delta_b^a = \begin{cases} 1 & a = b \\ 0 & a \neq b \end{cases} \tag{3}$$

Let's have a look at the transformation expression

$$\begin{aligned} \delta_b'^a &= \frac{\partial x'^a}{\partial x^i} \frac{\partial x^j}{\partial x'^b} \delta_j^i && (4) \\ &= \frac{\partial x'^a}{\partial x^i} \frac{\partial x^i}{\partial x'^b} && (5) \\ &= \frac{\partial x'^a}{\partial x'^b} && (6) \\ &= \delta_b^a && (7) \end{aligned}$$

The third line follows from the second by recognizing that the second line is the chain rule expression for the derivative in the third line. The last line follows from the fact that all the x' coordinates are independent of each other, so the derivative of any of them with respect to any other coordinate is zero, while the derivative of a coordinate with respect to itself is 1.

Thus not only does δ_b^a transform as a tensor, but it also has the same numerical value in all coordinate systems.

Note that this derivation does *not* work if we try to make delta to be either a pure contravariant or pure covariant tensor. If we start by defining (in some unprimed coordinate system):

$$\delta^{ij} = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases} \tag{8}$$

then applying the tensor transformation rule, we get

$$\delta'^{ab} = \frac{\partial x'^a}{\partial x^i}\frac{\partial x'^b}{\partial x^j}\delta^{ij} \tag{9}$$

$$= \sum_i \frac{\partial x'^a}{\partial x^i}\frac{\partial x'^b}{\partial x^i} \tag{10}$$

Although the index i is repeated, both occurrences are as a lower index so we can't use the summation convention, hence the explicit summation sign. However, this expression does not represent the chain rule derivative of anything in particular, and can't be simplified further. The expression depends on the coordinate systems, so this delta is not a numerical invariant, and doesn't transform like a tensor.

However, since δ_b^a is a tensor, we can raise or lower its indices using the metric tensor in the usual way. That is, we *can* get a version of δ with both indices raised or lowered, as follows:

$$\delta^{ab} = g^{cb}\delta_c^a = g^{ab} \tag{11}$$

$$\delta_{ab} = g_{ac}\delta_b^c = g_{ab} \tag{12}$$

In this sense, δ^{ab} and δ_{ab} are the upper and lower versions of the metric tensor. However, they can't really be considered versions of the Kronecker delta any more, as they don't necessarily satisfy 8. In other words, the only version of δ that is both a Kronecker delta *and* a tensor is the version with one upper and one lower index: δ_b^a.

PINGBACKS

Pingback: Tensor arithmetic

Pingback: Covariant derivative and connections

HIGHER ORDER DERIVATIVES ARE NOT TENSORS

Link to: physicspages home page.

To leave a comment or report an error, please use the auxiliary blog.

Reference: d'Inverno, Ray, *Introducing Einstein's Relativity* (1992), Oxford Uni Press. - Section 5.6 and Problem 5.9.

We've seen that the tangent to a curve is a contravariant tensor (vector, actually), since it transforms according to

$$\frac{dx'^a}{du}=\frac{\partial x'^a}{\partial x^i}\frac{dx^i}{du} \tag{1}$$

Also, the first derivative of a function is a covariant tensor, as it transforms according to

$$\frac{\partial g}{\partial x'^a}=\frac{\partial x^i}{\partial x'^a}\frac{\partial g}{\partial x^i} \tag{2}$$

You might think that higher order derivatives are also tensors, but this turns out not to be the case. If we take the derivative of the last equation with respect to another of the primed coordinates x'^c, we get

$$\frac{\partial^2 g}{\partial x'^a\partial x'^c} = \frac{\partial^2 x^i}{\partial x'^a\partial x'^c}\frac{\partial g}{\partial x^i}+\frac{\partial x^i}{\partial x'^a}\frac{\partial^2 g}{\partial x^i\partial x'^c} \tag{3}$$

$$= \frac{\partial^2 x^i}{\partial x'^a\partial x'^c}\frac{\partial g}{\partial x^i}+\frac{\partial x^i}{\partial x'^a}\frac{\partial x^j}{\partial x'^c}\frac{\partial^2 g}{\partial x^i\partial x^j} \tag{4}$$

where we used the chain rule on the second term. In order for $\frac{\partial^2 g}{\partial x^i\partial x^j}$ to transform like a tensor, the first term in the last line would have to be zero, but it's clearly not, in general.

Higher order derivatives will also have extraneous terms, so it is only first derivatives that behave as tensors.

PINGBACKS

Pingback: Electromagnetic field tensor: justification

Pingback: Christoffel symbols

Pingback: Lie derivatives

TENSOR ARITHMETIC

Link to: physicspages home page.

To leave a comment or report an error, please use the auxiliary blog.

Reference: d'Inverno, Ray, *Introducing Einstein's Relativity* (1992), Oxford Uni Press. - Section 5.8 and Problems 5.10 - 5.14.

Since tensors are generalizations of vectors, it's not surprising that many of the arithmetic properties of vectors that you may be familiar with also apply to tensors.

We've seen that the tangent to a curve is a contravariant tensor (vector, actually), since it transforms according to

$$\frac{dx'^a}{du}=\frac{\partial x'^a}{\partial x^i}\frac{dx^i}{du} \tag{1}$$

Also, the first derivative of a function is a covariant tensor, as it transforms according to

$$\frac{\partial g}{\partial x'^a}=\frac{\partial x^i}{\partial x'^a}\frac{\partial g}{\partial x^i} \tag{2}$$

Because these transformation rules are linear, the linear arithmetic operations of addition, subtraction and multiplication by a scalar, when applied to tensors, yield other tensors.

For example, if we have two tensors Y^a_{bc} and Z^a_{bc}, their sum is also a tensor of the same type, which we can call X^a_{bc}:

$$X'^a_{bc} = Y'^a_{bc}+Z'^a_{bc} \tag{3}$$

$$= \frac{\partial x'^a}{\partial x^i}\frac{\partial x^j}{\partial x'^b}\frac{\partial x^k}{\partial x'^c}Y^i_{\ jk}+\frac{\partial x'^a}{\partial x^i}\frac{\partial x^j}{\partial x'^b}\frac{\partial x^k}{\partial x'^c}Z^i_{\ jk} \tag{4}$$

$$= \frac{\partial x'^a}{\partial x^i}\frac{\partial x^j}{\partial x'^b}\frac{\partial x^k}{\partial x'^c}\left(Y^i_{\ jk}+Z^i_{\ jk}\right) \tag{5}$$

$$= \frac{\partial x'^a}{\partial x^i}\frac{\partial x^j}{\partial x'^b}\frac{\partial x^k}{\partial x'^c}X^i_{\ jk} \tag{6}$$

Obviously the same argument works for the difference of two tensors.

Note that it makes sense to define addition and subtraction only between tensors of the same type. In elementary linear algebra, for example, it makes no sense to talk about the sum or difference of a scalar and a vector, since the two objects are fundamentally different and cannot be combined in this way.

A second rank tensor such as X^{ab} is symmetric if $X^{ab} = X^{ba}$ for all pairs of indexes. A tensor is anti-symmetric if $X^{ab} = -X^{ba}$. Symmetry and anti-symmetry are examples of *tensorial properties*, which are properties that, if true in one coordinate system, are true in all coordinate systems. For symmetry, we have

$$X'^{ab} = \frac{\partial x'^a}{\partial x^i}\frac{\partial x'^b}{\partial x^j}X^{ij} \tag{7}$$

$$= \frac{\partial x'^a}{\partial x^i}\frac{\partial x'^b}{\partial x^j}X^{ji} \tag{8}$$

$$= X'^{ba} \tag{9}$$

The argument for anti-symmetry is the same, except a minus sign is introduced in the second line.

If X^{ab} is anti-symmetric and Y_{ab} is symmetric, then their product (with implied summation) is

$$X^{ab}Y_{ab} = -X^{ba}Y_{ba} \tag{10}$$

$$= -X^{ab}Y_{ab} \tag{11}$$

The first line follows from the anti-symmetry of X^{ab}. The second line follows from the fact that a and b are dummy indexes so it doesn't matter what we call them, so we are justified in just swapping them around. Thus the equation has the form $X^{ab}Y_{ab} = -X^{ab}Y_{ab}$, which means that $X^{ab}Y_{ab} = 0$. This is because the terms in the sum cancel each other in pairs. (Note that for an anti-symmetric tensor, $X^{aa} = -X^{aa} = 0$.)

A tensor can be split into a sum of symmetric and anti-symmetric parts. For a rank 2 tensor, clearly the following tensor is symmetric:

$$X^{(ab)} \equiv \frac{1}{2}\left(X^{ab} + X^{ba}\right) \tag{12}$$

and the following tensor is anti-symmetric:

$$X^{[ab]} \equiv \frac{1}{2}\left(X^{ab} - X^{ba}\right) \tag{13}$$

The special bracket notation is used to denote symmetric and anti-symmetric tensors.

The sum gives us back the original tensor:

$$X^{ab} = X^{(ab)} + X^{[ab]} \tag{14}$$

The notion of symmetry can be extended to tensors of arbitrary rank:

$$X_{(a_1 a_2 \ldots a_r)} = \frac{1}{r!}\sum(\text{all permutations of the indexes}) \tag{15}$$

$$X_{[a_1 a_2 \ldots a_r]} = \frac{1}{r!}\left[\sum(\text{all even permutations}) - \sum(\text{all odd permutations})\right] \tag{16}$$

An even permutation is one where an even number of swaps is needed to get from the original order to a given order, and an odd permutation requires an odd number of swaps. The even permutations of the starting index order *abc* are *cab* and *bca*, leaving *bac*, *cba* and *acb* as the odd permutations. A symmetric tensor is one where all components with a particular set of indexes, permuted in any order, are equal.

A tensor can be *contracted* by setting a pair of one upper and one lower index equal, giving a summation. For example, given a rank 3 tensor X^a_{bc} we can form the contraction $Y_c = X^a_{ac}$. The result of a contraction is another tensor, as we can verify by finding its transformation:

$$\begin{aligned}
Y'_c &= X'^a_{ac} && (17)\\
&= \frac{\partial x'^a}{\partial x^i}\frac{\partial x^j}{\partial x'^a}\frac{\partial x^k}{\partial x'^c}X^i_{\ jk} && (18)\\
&= \frac{\partial x^j}{\partial x^i}\frac{\partial x^k}{\partial x'^c}X^i_{\ jk} && (19)\\
&= \delta^j_i\frac{\partial x^k}{\partial x'^c}X^i_{\ jk} && (20)\\
&= \frac{\partial x^k}{\partial x'^c}X^i_{ik} && (21)\\
&= \frac{\partial x^k}{\partial x'^c}Y_k && (22)
\end{aligned}$$

Thus the contracted tensor transforms as a covariant vector.

We've seen that the Kronecker delta is a tensor if its indexes are written as δ^a_b. If we contract this tensor, we get (in an n-dimensional manifold)

$$\delta^a_a = \sum_{a=1}^{n} 1 = n \tag{23}$$

How about $\delta^a_b\delta^b_a$? This is easiest to see by writing out the sums.

$$\delta_b^a \delta_a^b = \sum_{a=1}^{n} \sum_{b=1}^{n} \delta_b^a \delta_a^b \tag{24}$$

$$= \sum_{a=1}^{n} \delta_a^a \delta_a^a \tag{25}$$

$$= \sum_{a=1}^{n} 1 \times 1 \tag{26}$$

$$= n \tag{27}$$

The second line follows since $\delta_b^a \neq 0$ only when $b = a$.

PINGBACKS

Pingback: Congruence of curves

LIE BRACKETS (COMMUTATORS)

Link to: physicspages home page.

To leave a comment or report an error, please use the auxiliary blog.

Reference: d'Inverno, Ray, *Introducing Einstein's Relativity* (1992), Oxford Uni Press. - Section 5.9 and Problems 5.15, 5.16 (v).

When we began looking at quantum mechanics, we encountered the commutator of two operators, defined as

$$[A,B] \equiv AB - BA \tag{1}$$

In quantum mechanics, some operators (the most famous being the position and momentum operators) do not commute, and in fact, the generalized uncertainty principle says that only operators that commute can be measured simultaneously with arbitrary precision.

In tensor analysis, we've seen that the tangent vector field to a manifold can be written as the operator

$$X = X^a \partial_a \tag{2}$$

Since this operator involves derivatives, we might expect that the commutator of two such operators would be non-zero (since that's what happend with the position and momentum operators in quantum mechanics). The commutator of two vector fields is also known as a *Lie bracket*, (where 'Lie' is pronounced 'lee') but is defined in the same way as in quantum mechanics.

The commutator of two vector fields is again a vector field, as can be verified by direct calculation. As always with operators involving derivatives, we need a dummy function f on which to operate, so we get

$$[X,Y]f = X^a \partial_a \left(Y^b \partial_b f\right) - Y^a \partial_a \left(X^b \partial_b f\right) \tag{3}$$

$$= X^a \left(\partial_a Y^b\right)(\partial_b f) + X^a Y^b \partial^2_{ab} f - Y^a \left(\partial_a X^b\right)(\partial_b f) - Y^a X^b \partial^2_{ab} f \tag{4}$$

$$= X^a \left(\partial_a Y^b\right)(\partial_b f) - Y^a \left(\partial_a X^b\right)(\partial_b f) \tag{5}$$

Removing the dummy function, we get

$$[X,Y]=X^{a}\left(\partial_{a}Y^{b}\right)\partial_{b}-Y^{a}\left(\partial_{a}X^{b}\right)\partial_{b} \tag{6}$$

which is a vector field with components

$$[X,Y]^{b}=X^{a}\partial_{a}Y^{b}-Y^{a}\partial_{a}X^{b} \tag{7}$$

It's obvious from the definition that

$$[X,X] = 0 \tag{8}$$
$$[X,Y] = -[Y,X] \tag{9}$$

There is a third identity known as *Jacobi's identity* that is less obvious:

$$[X,[Y,Z]]+[Z,[X,Y]]+[Y,[Z,X]]=0 \tag{10}$$

This is true for commutators in general, and not just for vector fields as defined here. It can be proved by writing out the terms.

$$\begin{aligned}
[X,[Y,Z]] &= XYZ-XZY-YZX+ZYX\\
[Z,[X,Y]] &= ZXY-ZYX-XYZ+YXZ\\
[Y,[Z,X]] &= YZX-YXZ-ZXY+XZY
\end{aligned}$$

Adding up the right hand side, we see that the terms cancel in pairs.

As an example of Lie brackets, we can look at the operators X, Y and Z that we used in the post on tangent space. In rectangular coordinates, these operators are

$$X = \partial_{x} \tag{11}$$
$$Y = \partial_{y} \tag{12}$$
$$Z = -y\partial_{x}+x\partial_{y} \tag{13}$$

To work out the commutators, we can use equation 7 above. For that, we need the components of the vectors, which are $X^{a}=(1,0)$, $Y^{a}=(0,1)$ and $Z^{a}=(-y,x)$.

$$[X,Y]^b = X^a\partial_a Y^b - Y^a\partial_a X^b \tag{14}$$
$$= (0,0) \tag{15}$$
$$[X,Z]^1 = X^a\partial_a Z^1 - Z^a\partial_a X^1 \tag{16}$$
$$= \partial_x(-y) + 0 - (0+0) \tag{17}$$
$$= 0 \tag{18}$$
$$[X,Z]^2 = X^a\partial_a Z^2 - Z^a\partial_a X^2 \tag{19}$$
$$= \partial_x x + 0 - (0+0) \tag{20}$$
$$= 1 \tag{21}$$
$$[Y,Z]^1 = Y^a\partial_a Z^1 - Z^a\partial_a Y^1 \tag{22}$$
$$= 0 + \partial_y(-y) - (0+0) \tag{23}$$
$$= -1 \tag{24}$$
$$[Y,Z]^2 = Y^a\partial_a Z^2 - Z^a\partial_a Y^2 \tag{25}$$
$$= 0 + \partial_y x - (0+0) \tag{26}$$
$$= 0 \tag{27}$$

Thus the commutator operators are

$$[X,Y] = 0 \tag{28}$$
$$[X,Z] = \partial_y = Y \tag{29}$$
$$[Y,Z] = -\partial_x = -X \tag{30}$$

PINGBACKS

Pingback: Riemann tensor and covariant contraction

TANGENT SPACE

Link to: physicspages home page.

To leave a comment or report an error, please use the auxiliary blog.

Reference: d'Inverno, Ray, *Introducing Einstein's Relativity* (1992), Oxford Uni Press. - Section 5.9 and Problems 5.16 (i - iv).

When we looked at contravariant vectors, we examined the case of the directional derivative along a curve, and showed that if we have the curve defined parametrically by

$$f=f(x^{a}(u)) \tag{1}$$

where u is the parameter, then the directional derivative is given by (using the chain rule):

$$\frac{df}{du}=\frac{\partial f}{\partial x^{i}}\frac{dx^{i}}{du} \tag{2}$$

In 3-d Euclidean geometry, the derivative df/du is the tangent to the curve. By drawing all possible curves through a particular point P and taking the directional derivative of each curve, we can generate a collection of tangents called the *tangent space*. In 3-d Euclidean geometry, the tangent space can be visualized as the plane tangent to a surface at P. In higher dimensions and in non-Euclidean geometries, the concept is not so easy to visualize, but the mathematics generalizes in a straightforward way.

It's important to note that in non-Euclidean geometries, the space in which the tangent space lies may be different from that of the space which it is tangent to. Examples of this must be delayed until we've looked at such non-Euclidean spaces.

We can rewrite the above equation as an operator equation:

$$\frac{df}{du}=Xf \tag{3}$$

where

$$X\equiv\frac{dx^{i}}{du}\frac{\partial}{\partial x^{i}} \tag{4}$$

As a shorthand notation, since the partial derivative turns up often, we can write

$$\partial_i \equiv \frac{\partial}{\partial x^i} \tag{5}$$

$$X = \frac{dx^i}{du}\partial_i \tag{6}$$

As a further shorthand, we can define the components of the contravariant vector to be $X^a \equiv dx^a/du$, so we get

$$X = X^a \partial_a \tag{7}$$

This operator is invariant under a change of coordinates, as can be shown fairly easily. Since the quantities X^a make up a contravariant vector, we know how they transform, so we get

$$X'^a \partial'_a = \frac{\partial x'^a}{\partial x^i} X^i \partial'_a \tag{8}$$

$$= \frac{\partial x'^a}{\partial x^i} X^i \frac{\partial x^j}{\partial x'^a} \partial_j \tag{9}$$

$$= \delta_i^j X^i \partial_j \tag{10}$$

$$= X^i \partial_i \tag{11}$$

As an example, suppose we have a vector field in two dimensions given in rectangular coordinates as $X^a = (1,0)$. That is, the vector field 'points' in the x direction and has a constant magnitude over the entire plane. To transform to polar coordinates, we can use the contravariant transformation

$$X'^a = \frac{\partial x'^a}{\partial x^i} X^i \tag{12}$$

For polar coordinates $r = \sqrt{x^2 + y^2}$ and $\theta = \tan^{-1}(y/x)$ we get

$$X'^1 = \frac{\partial r}{\partial x} X^1 + 0 \tag{13}$$
$$= \frac{x}{r} \tag{14}$$
$$= \cos\theta \tag{15}$$
$$X'^2 = \frac{\partial \theta}{\partial x} X^1 + 0 \tag{16}$$
$$= -\frac{\sin\theta}{r} \tag{17}$$

The gradient operator can be written in the two coordinate systems as

$$\nabla f = \frac{\partial f}{\partial x}\hat{\mathbf{i}} + \frac{\partial f}{\partial y}\hat{\mathbf{j}} \tag{18}$$
$$= \frac{\partial f}{\partial r}\hat{\mathbf{r}} + \frac{\partial f}{\partial \theta}\frac{\hat{\boldsymbol{\theta}}}{r} \tag{19}$$

The relation between the unit vectors is

$$\hat{\mathbf{r}} = \cos\theta\hat{\mathbf{i}} + \sin\theta\hat{\mathbf{j}} \tag{20}$$
$$\hat{\theta} = -\sin\theta\hat{\mathbf{i}} + \cos\theta\hat{\mathbf{j}} \tag{21}$$

If we take the dot product of ∇f with each unit vector, we can get relations between the derivatives. We get

$$\nabla f \cdot \hat{\mathbf{r}} = \cos\theta\partial_x f + \sin\theta\partial_y f \tag{22}$$
$$= \partial_r f \tag{23}$$
$$\nabla f \cdot \hat{\theta} = -\sin\theta\partial_x f + \cos\theta\partial_y f \tag{24}$$
$$= \frac{1}{r}\partial_\theta f \tag{25}$$
$$\nabla f \cdot \hat{\mathbf{i}} = \cos\theta\partial_r f - \frac{\sin\theta}{r}\partial_\theta f \tag{26}$$
$$= \partial_x f \tag{27}$$
$$\nabla f \cdot \hat{\mathbf{j}} = \sin\theta\partial_r f - \frac{\cos\theta}{r}\partial_\theta f \tag{28}$$
$$= \partial_y f \tag{29}$$

In rectangular coordinates, the operator X is

$$X = X^a\partial_a \tag{30}$$
$$= \partial_x \tag{31}$$

In polar coordinates, we get

$$X' = X'^a\partial'_a \tag{32}$$
$$= \cos\theta\,\partial_r - \frac{\sin\theta}{r}\partial_\theta \tag{33}$$
$$= \cos^2\theta\,\partial_x + \cos\theta\sin\theta\,\partial_y + \sin^2\theta\,\partial_x - \cos\theta\sin\theta\,\partial_y \tag{34}$$
$$= \partial_x \tag{35}$$

Thus the operator is indeed the same in the two coordinate systems.

For a second vector field $Y^a = (0,1)$ we can do the same analysis to find that

$$Y'^a = \left(\sin\theta, \frac{\cos\theta}{r}\right) \tag{36}$$
$$Y = \partial_y \tag{37}$$
$$= \sin\theta\,\partial_r + \frac{\cos\theta}{r}\partial_\theta \tag{38}$$

Finally, if $Z^a = (-y,x)$ we get

$$Z'^a = \left(-y\cos\theta + x\sin\theta, y\frac{\sin\theta}{r} + x\frac{\cos\theta}{r}\right) \tag{39}$$
$$= (0,1) \tag{40}$$
$$Z = -y\partial_x + x\partial_y \tag{41}$$
$$= \partial_\theta \tag{42}$$

PINGBACKS

Pingback: Lie brackets (commutators)

Pingback: Tangent space: partial derivatives as basis vectors