Lesson 2: Tensor mathematics

Notes from Prof. Susskind video lectures publicly available on YouTube

Introduction

A good notation, as we said, will carry you a long way. When it is well conceived, it just sort of automatically tells you what to do next. That means that you can do physics in a completely mindless way.

It is like having Tinkertoys. It is pretty clear where the stick has to go into. It has to go into the thing with the hole. You can try putting a hole into a hole or forcing a stick into a stick. There is only one thing you can do. You can put the stick into the hole, and the other end of the stick can go into another hole. Then there are more sticks and more holes you can put them into, etc.

The notation of general relativity is much like that. If you follow the rules, you almost can't make mistakes. But you have to learn the rules. They are the rules of tensor algebra and tensor analysis.

Flat space

The question we are aiming at in this chapter is to understand enough about tensor algebra and analysis, and metrics, to be able to distinguish a flat geometry from a non-flat geometry. That seems awfully simple. Flat means like a plane. Non-flat means with bumps and lumps in it. And you would think we could tell the difference very easily. Yet sometimes it is not so easy.

For example as discussed in last chapter, if I consider this page, it is flat. If I roll it or furl it, the page now looks curved but it is not really curved. It is exactly the same page. The relationship between the parts of the page, the distances between the letters, the angles, and so forth, don't change. At least the distances between the letters *measured along the page* don't change. So a folded page, if we don't stretch it, if we don't modify the relations between its parts, doesn't acquire a curvature.

Technically it introduces what is called an *extrinsic curvature*. Extrinsic curvature has to do with the way a space – in this case the page – is embedded in a higher-dimensional space. For instance whatever I do with the page is embedded in the three dimensional space of the room. When the page is laid out flat on the desk, it is embedded in the embedding space in one way. When it is furled like in figure 1, it is embedded in the same space in another way.

Figure 1: Intrinsic and extrinsic geometries and curvatures: the intrinsic geometry of the page remains flat.

The extrinsic curvature which we perceive has to do with *how* the space of the page is embedded in the larger space. But it has nothing to do with its *intrinsic geometry*.

If you like, you can think of the intrinsic geometry as the geometry of a tiny little bug that moves along the surface. It cannot look out of the surface. It only looks around while crawling along the surface. It may have surveying instruments with which it can measure distances along the surface. It can draw a triangle, measure also the angles within the surface, and do all kinds of interesting geometric studies. But it never sees the surface as embedded in a larger space.

Consequently the bug will never detect that the page might be embedded in different ways in a higher dimensional space. It will never detect it if we create a furl like in figure 1, or if we remove the furl and flatten the page out again. The bug just learns about the intrinsic geometry.

The intrinsic geometry of the surface means the geometry that is independent of the way the surface is embedded in a larger space.

General relativity and Riemannian geometry, and a lot of other geometries, are all about the intrinsic properties of the geometry of the space under consideration. It doesn't have to be two dimensional. It can have any number of dimensions.

Another way to think about the intrinsic geometry of a space is this. Imagine sprinkling a bunch of points on this page – or on a three dimensional space, but then we would have to fiddle with it in four dimensions or more... Then draw lines between them so they triangulate the space. And then state what the distance between every pair of neighboring points is. *Specifying those distances specifies the geometry.*

Sometimes that geometry can be flattened out without changing the length of any of these little links. In the case of a two-dimensional surface, it means laying it out flat on the desk without stretching it, tearing it, or creating any distorsion. Any small equilateral triangle has to remain an equilateral triangle. Every small little square has to remain a square, etc.

But if the surface is intrinsically non-flat there will be small constructions that cannot be flattened out. The other day on his motorbike the second author saw on the road the following bulge, probably due to pine roots, with lines drawn on it, and a warning painted on the pavement.

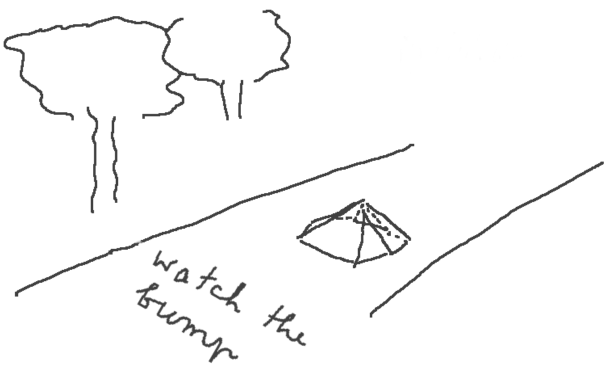

Figure 2: Watch the bump.

The road menders must have taken a course in general relativity! Such a bump cannot be flattened out without stretching or compressing some distances.

A curved space is basically one which cannot be flattened out without distorting it. It is an intrinsic property of the space, not extrinsic.

Metric tensor

We want to answer the mathematical question: given a space and its metric defined by the following equation

$$dS^2 = g_{mn}(X)\ dX^m dX^n \tag{1}$$

is it really flat or not?

It is important to understand that the space may be intrinsically curved, like the road with a bump in figure (2), or we *may think* that it is curved because equation (1) looks complicated, when actually it is intrinsically flat.

For instance we can draw on a flat page a bunch of funny curvilinear coordinates as in figure 3. Now let's forget that we look comfortably at the page from our embedding 3D Euclidean space. At first sight the coordinate axes X's suggest that it is curved.

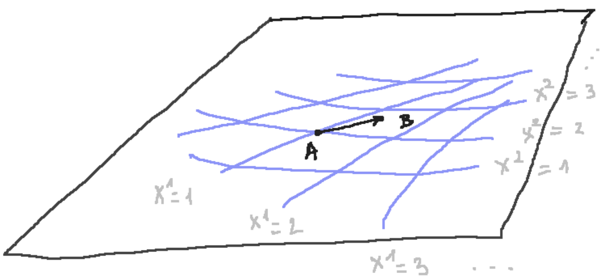

Figure 3: Curvilinear coordinates X's of a flat page.

At each point A, if we want to compute the distance between A and a neighboring point B, we cannot apply Pythagoras theorem. We have to apply Al-Kashi theorem which generalizes Pythagoras taking into account the cosine of the angle between the coordinate axes. And also perhaps we have to correct for units which are not unit distances on the axes.

Yet the page is intrinsically flat, be it rolled or not in the embedding 3D Euclidean space. It is easy to find a set of coordinates Y's which will transform equation (1) into Pythagoras theorem. On the pages of school notebooks they are even usually shown. And it doesn't disturb us to look at them, interpret them, and use them to locate a point, even when the page is furled.

Our mathematical goal matches closely the question we addressed in the last chapter of whether there is a real gravitational field or the apparent gravitational field is just due to an artefact of funny space-time coordinates. For instance in figure (4) of chapter 1 the curvilinear coordinates were due to the accelerated frame we were using, not to tidal forces. The space-time was intrinsically flat.

So we want to tackle the mathematical question. Typically we are given the metric tensor of equation (1). The mathematical question is a hard one. It will keep us busy during the entire chapter and more.

Before we come to it, we need to get better acquainted with tensors. We have begun to talk about them in the last chapter. We introduced the basic contravariant and covariant transformation rules. In this chapter, we want to give a more formal presentation of tensors.

Scalars and vectors are special cases of tensors. Tensors are the general category of objects we are interested in.

Scalar, vector and tensor fields

So for us, tensors are collections of indexed values which depend on coordinate systems. And they transform according to certain rules when we go from one coordinate system to another.

We are going to be interested in spaces such that at every point P in space, located by its coordinates X in some coordinate system, there may be some quantities associated with that point – what we call fields. And those quantities will be tensors. There will also be all kinds of quantities that will not be tensors. But in particular we will be interested in tensor fields.

The simplest kind of tensor field is a scalar field $S(X)$. A scalar field is a function which to every point of space associates a number – a scalar –, and everybody, no matter what coordinate system he or she uses, agrees on the value of that scalar. So the transformation properties in going, let's say, from the X^m coordinates to the Y^m coordinates is simply that the value of S at a given point P doesn't change.

We could use extremely heavy notations to express this fact in the most unambiguous way. But we will simply denote it

$$S'(Y) = S(X) \tag{2}$$

The right hand side and the left hand side denote *the value of the same field at the same point* P, one in the Y system, the other in the X system. Y is the coordinates of P in the Y system, X is the coordinates of P in the X system. And we add a prime to S when we talk of its value at P using the Y coordinates. With practice, equation (2) will become clear and unambiguous.

To understand what distinguishes a scalar field from any old scalar function, notice that if we fix the coordinate system then "the first coordinate of a vector field" is a scalar function, but is not a scalar field, because it depends on the coordinate system, and it changes if we change coordinate system.

Let's represent, on a two-dimensional variety, the X coordinate system. And now, to avoid confusion, let's not embed the surface in any larger Euclidean space.

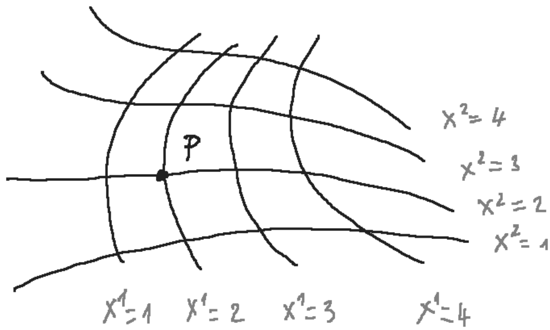

Figure 4: Curvilinear coordinates X's
on a two-dimensional curved variety.

Any point P of the space is located by the values its coordinates X^1 and X^2. Of course we could think of a higher dimensional variety. There would then be more coordinates.

Globally, we denote them X^m.

Now on the same space, there could be another coordinate system, Y, to locate points.

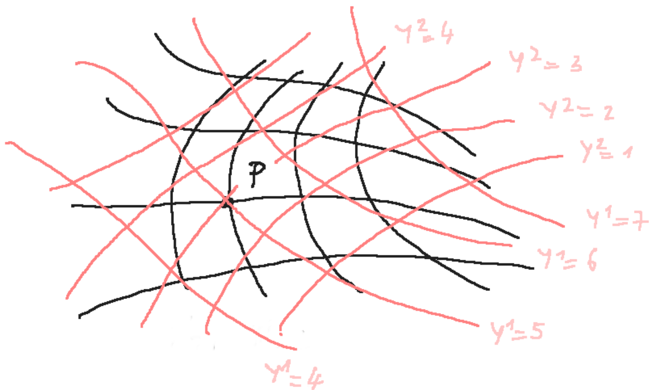

Figure 5: Second coordinate system Y on a two-dimensional curved variety.

In our figure, the point P has coordinates $(2,\ 2)$ in the X system, and $(5,\ 3)$ in the Y system. Of course these coordinates don't have to be integers. They can take their values in the set of real numbers, or even in other sets.

What is important to note is that at any point P, there are two collections of coordinates

$$X^m \text{ and } Y^m$$

The X^m and Y^m are related. At any point P, each coordinate X^m is a function of all the Y^m. And conversely. We write it this way

$$X^m = X^m(Y) \tag{3a}$$

and

$$Y^m = Y^m(X) \tag{3b}$$

This is a coordinate transformation of some kind, and its inverse. It can be pretty complicated. We will only assume that functions (3a) and (3b) are continuous, and that we can differentiate them when needed. But there is nothing more special than that.

Scalars transform trivially. If you know the value of S at a point P, you know it no matter what coordinate system you use.

Next are vectors. They come in two varieties: contravariant vectors which we denote with an upstairs index

$$V^m$$

and covariant vectors with a dowstairs index

$$V_m$$

We spoke about them in the last chapter. Now we are going to see a little bit about their geometrical interpretation. What it intuitively means to be contravariant or to be covariant?

Geometric interpretation of contravariant and covariant components of a vector

Let's consider a coordinate system, and draw its axes as straight lines because we are not interested at the moment in the fact that the coordinates may be curved and vary in direction from place to place. We could also think of them *locally*, where every variety is approximately flat (a surface, locally, is like a plane) and every coordinate system locally is formed of approximatelty straight lines or surfaces if we are in more than two dimensions.

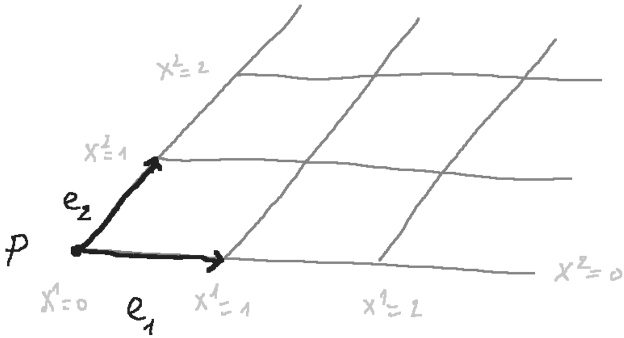

Figure 6: Coordinate system at point P.

We are mostly concerned with the fact that the coordinate axes may not be perpendicular, and with what the implications of the non perpendicularity of these coordinates are. Furthermore the distance between two axes, say $X^1 = 0$ and $X^1 = 1$, is not necessarily 1. The values of the coordinates are just numerical labels, which don't correspond directly to distances.

Now let's introduce some vectors. On our two-dimensional variety, we introduce two

$$e_1 \text{ and } e_2$$

as shown on figure 6.

If we had three dimensions, there would be a third vector e_3 sticking out of the page, possibly slanted. We can label these vectors

$$e_i$$

As i goes from 1 to the number of dimensions, the geometric vectors e_i's correspond to the various directions of the coordinate system.

Next step in the geometric explanation of contravariant and covariant vectors: we consider an arbitrary vector V, see figure 7.

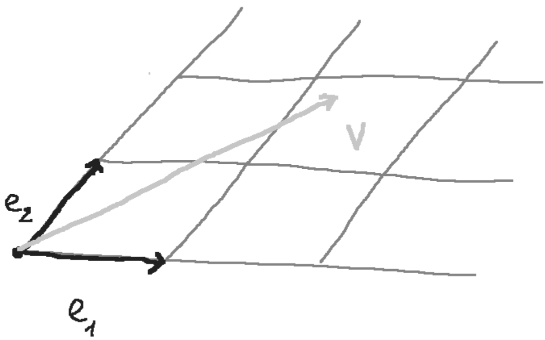

Figure 7: Vector V.

The vector V can be expanded into a linear combination of the e_i's. We shall write V^i for the i-th coefficient, and suppose there are 3 dimensions. Then

$$V = V^1 e_1 + V^2 e_2 + V^3 e_3 \tag{4}$$

The things which are the vectors, on the right hand side of this formula, are the e_i's. The V^i's are actually numbers. They are the components of the vector V in the e_i basis.

The coefficients V^i are called the contravariant components of the vector V. It is just a name. And there is nothing in what I did that required me to put the index 1 of e_1 downstairs and not upstairs, and the index 1 of V^1 upstairs. It is a convention to write the expansion of V in the form of equation (4).

So, first of all, we see what the contravariant components are. They are the expansion coefficients of V, that is, the numbers that we have to put in front of the three vectors e_1, e_2 and e_3 to express a given vector as a sum of vectors colinear to the basis. This jives with what we have said previously: ordinary vectors are contravariant vectors.

Next step: we look at the projection of V on the e_i's using the dot product. Let's start with e_1

$$V.e_1$$

Now if we were just using conventional Cartesian coordinates, perpendicular to each other, and if the e_i's really were unit vectors, that is, if the distance representing each coordinate separation was one unit of whatever the units we are dealing with, then the coefficients V^1, V^2 and V^3 would be the same as the dot products. For instance we would have $V.e_1$ equal the first contravariant component of V.

However, when we have a peculiar coordinate system with angles and with non-unit separations between the successive coordinate lines in figure 7, this is not true. So let's see if we can work out V with the values $V.e_1$, $V.e_2$, $V.e_3$...

Incidentally, $V.e_1$ *is called* V_1*, denoted with a covariant index.*

Notice how notations fit nicely together. We can write equation (4) as

$$V = V^m e_m \tag{5}$$

using the Einstein summation convention.

Now let's see how we can relate the contravariant components V^m and the covariant components V_n. To reach that goal we take the dot product of each side of equation (5) with e_n. We get

$$V.e_n = V^m e_m.e_n \tag{6}$$

And $V.e_n$ is by definition V_n.

$e_m.e_n$ is something new. Let's isolate it. It has two lower indices. We will see that it turns out to be the metric tensor (expressed in the e_i's basis).

Let's see this connection between $e_m.e_n$ and the metric tensor. The length of a vector is the dot product of the vector with itself. Let's calculate the length of V. Using twice equation (5) we have

$$V.V = V^m e_m.V^n e_n \tag{7}$$

We must use two different indices m and n. Recall indeed that, in the implicit summation formula $V^m e_m$, the symbol m is only a dummy index. So in order not to mix things up, we use another dummy index n for the second expression of V.

If you are not yet totally at ease with Einstein summation convention, remember that, written explicitely, the right hand side of equation (7) means nothing more than

$$(V^1 e_1 + V^2 e_2 + V^3 e_3).(V^1 e_1 + V^2 e_2 + V^3 e_3)$$

But now the right hand side of equation (7) can also be reorganized as

$$V.V = V^m V^n (e_m.e_n) \tag{8}$$

The quantity $e_m.e_n$ we call g_{mn}. So equation (8) rewrites

$$V.V = V^m V^n g_{mn} \tag{9}$$

This is characteristic of the metric tensor. It tells you how to compute the length of a vector.

The vector could be for instance a small displacement dX. Then equation (9) would be the computation of the length of a little interval between two neighboring points

$$dX.dX = dX^m dX^n g_{mn} \tag{10}$$

So now we have a better understanding of the difference between covariant and contravariant indices, that is to say covariant and contravariant components. Contravariant components are the coefficients we use to construct a vector V out of the basis vectors. Covariant components are the dot products of V with the basis vectors. They are different geometric things. They would, however, be the same if we were talking about ordinary Cartesian coordinates.

We inserted that discussion in order to give the reader some geometric idea of what covariant and contravariant means and also what the metric tensor is. For a given collection of basis vectors e_i's and a given vector V, let's summarize all this in the following box

$$\boxed{\begin{aligned} V &= V^m e_m \\ V_n &= V.e_n \\ g_{mn} &= e_m.e_n \end{aligned} \qquad (11)}$$

These relations are very important, and we will make frequent use of them in the construction of the theory of general relativity.

Let's just make one more note about the case when the coordinates axes are Cartesian coordinates. Then, as we saw, the contravariant and the covariant components of V are the same. And the metric tensor is the unit matrix. This means that the basis vectors are perpendicular *and* of unit length. Indeed, they could be orthogonal without being of unit length. In polar coordinates (see figure 14 of chapter 1, and figure 8 below), the basis vectors at any point P on the sphere are orthogonal, but they are not all of unit length. The longitudinal basis vector has a length which depends on the latitude. It is equal to the cosine of the latitude. That is why, on the sphere of radius one, to compute the square of the length of an element dS we can use Pythagoras theorem, but we must add $d\theta^2$ and $cos^2\theta\ d\phi^2$.

Also note that nothing enjoins us to represent the sphere in perspective, embedded in the usual 3D Euclidean space, like we did in figure 14 of chapter 1. We can also represent it – or part of it – on a page. Let's do it for a section of the Earth around one of its poles.

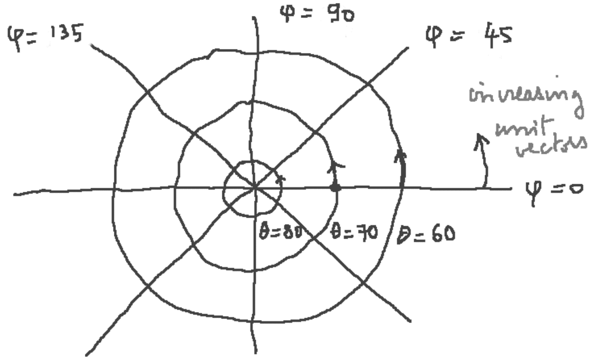

Figure 8: Map of the Earth around the North pole.

This is a representation on a page – therefore, out of necessity, flat – of a non-flat Riemannian surface with curvilinear coordinates, in this case a section of sphere in polar coordinates. As already mentioned, we touch here on the classical problem of cartographers: how to represent a section of sphere on a page, that is, how to make useful maps for mariners (see footnote on page 39 of chapter 1).

This ought to clarify the fact that we can represent on a page a curved, truly non-flat, variety, and a curvilinear coordinate system on it.

This is also what is achieved by ordnance survey maps, which can show hills and valleys, slopes, distances on inclined land, gradients and things like that, see figure 9. The curvy lines shown are the lines of equal height with respect to an underlying flat plane, which is a locally flat small section of the sphere[1] on which we represent the montainous relief. The grid of straight lines is a coordinate system on the sphere.

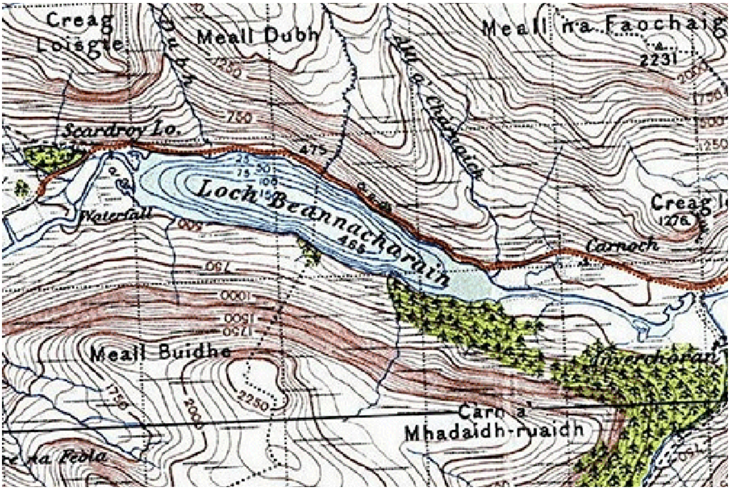

Figure 9: Ordnance survey map.

Since the notions of curved surfaces, and distances on them, and local curvatures are fundamental in general relativity, and we only treat them cursorily in this book, as groundwork for the physics, we advised the interested reader to go to any good simple manual on differential geometry oriented toward applications.

So now let's come to tensor mathematics.

[1] or more precisely the ellipsoid with which we represent the Earth

Tensor mathematics

Tensors are objects which are characterised by the way they transform under coordinate transformations. We already talked a little bit about them at the end of chapter 1. Now we want to go over again what we said and go further.

Notice that to say that tensors are characterized by the way they transform is no more strange than to say in $\mathbb{R}^3$ that (a, b, c) is a vector, or if you prefer "can be thought of as a vector", if and only if this collection of 3 numbers depends on a basis, is the expression of a thing in that basis, and it transforms in the usual way when we change basis. Let's go over this in more detail.

So let's start with a vector V. It has contravariant components in the X coordinates. We called them V^m. And it has contravariant components in the Y coordinates, which we called $(V')^m$.

In figure 7, if we change the coordinate system, keeping the abstract geometric vector V fixed in the space, we are still talking about the same vector, but we will clearly change its components. How do the contravariant components change when we change coordinates? We have seen the rule. Let us repeat it here. Remember, in the formula below, prime means "in the Y system" and unprimed means "in the X system".

$$(V')^m = \frac{\partial Y^m}{\partial X^n} V^n \tag{12}$$

And now let's look at a covariant vector. For the most typical example, we start from a scalar field $S(X)$ which we differentiate with respect to the X components, and the

Y components. We have seen the rule as well. The partial derivatives, which are covariant components, are related as follows

$$\frac{\partial S}{\partial Y^m} = \frac{\partial X^n}{\partial Y^m}\frac{\partial S}{\partial X^n} \tag{13}$$

Notice the difference. And notice how the notation carries you along. In equation (12)
- the index m is upstairs
- on the right hand side the proportionality factor is $\partial Y^m / \partial X^n$
- the sum is over n

Whereas in equation (13)
- the index m is downstairs
- on the right hand side the proportionality factor is $\partial X^n / \partial Y^m$
- the sum is still on n.

If there is no index n on the left hand side, but an index n appears on the right, then an index n upstairs has to be balanced by an index n downstairs. And we can "contract" them. This means that they represent a sum, are only dummy indices, and disappear. In both equations you can see the pattern. And as said, the notation pretty much carries you along.

Equation (12) is the standard form for the transformation property of contravariant components. And equation (13) is the standard form for the transformation property of covariant components, if they come from differentiating a scalar. More generally it would be equation (14) below

$$(W')_m = \frac{\partial X^n}{\partial Y^m} W_n \tag{14}$$

Let's go now to tensors of higher rank. A tensor of higher rank simply means a tensor with more indices. For the sake of pedagogy and completeness in this chapter 2, we overlap a bit what we did at the end of the last lesson.

We start with a tensor of rank two, with one contravariant index and one covariant index. It is nothing more than a "thing" represented in a given basis by a collection of numbers. These numbers are indexed with two indices. Furthermore in another basis the same "thing" is represented by another collection of numbers and the two collections are subject to specific transformation rules related to the relationship between the two bases. Let's consider the tensor in a Y basis, that is to say, a Y coordinate system. We denote it

$$(W')^m{}_n$$

The simplest example of such a thing would be, as we saw, just the product of two vectors, one with a contravariant index, one with a covariant index. By "product of the vectors" we mean the collection of all the products of components. What makes the thing a tensor is its transformation property. So let's write it

$$(W')^m{}_n = \frac{\partial Y^m}{\partial X^p} \frac{\partial X^q}{\partial Y^n} W^p{}_q \tag{15}$$

This tells us how a tensor of rank 2, with one contravariant and one covariant index, transforms. For each index on the left hand side, there must be a $\partial Y/\partial X$ or a $\partial X/\partial Y$ on the right hand side. And you simply track where the indices go.

Let's do another example of a tensor of rank 2 with two covariant indices

$$(W')_{mn}$$

how does it transform? By now you should begin to be able to write it mechanically

$$(W')_{mn} = \frac{\partial X^p}{\partial Y^m} \frac{\partial X^q}{\partial Y^n} W_{pq} \tag{16}$$

These rules are very general. If you take a tensor with any number of indices, the pattern is always the same. To express the transformation rules from an unprimed system X to a prime system Y, you introduce partial derivatives, in one sense or the other as we did, on the right hand side, and you sum over repeated indices.

We now know the basic notational device to express a tensor of any rank and type in one coordinate system or another.

Who invented it? Einstein was the one who dropped the summation symbol, because he realized he didn't need it. Gauss began to use formulas akin to equations (12) and (13) in his study of surfaces. Riemann continued in the development of non-Euclidean geometry. Ricci-Curbastro and Levi-Civita gave a formal presentation of tensor analysis in their fundamental work "Méthodes de calcul différentiel absolu et leurs applications", published in Mathematische Annalen, in March 1900.

The notation is the work of many, but it is very systematic.

Notice something about tensors. If they are zero in one frame, there are necessarily zero in any other too. This is obvious for scalars: if a scalar is 0 in one frame, it is 0 in every frame, because its value depend only on the geometric point where it is measured, not the coordinates of that point.

Now suppose a vector V is zero in some frame – let's say the X frame. To say that V is zero doesn't mean that some component is equal to zero, it means all of its components are equal zero. Then equation (12) or equation (14) show that it is going to be zero in any frame.

Likewise with any tensor, if all of its components are 0 in one frame, that is, in one coordinate system, then all of its components are 0 in every frame.

As a consequence, once we have written down an equation equating two tensors in one frame, for instance

$$T^{lmn}_{pqr} = U^{lmn}_{pqr}$$

it can be rewritten

$$T^{lmn}_{pqr} - U^{lmn}_{pqr} = 0$$

So, considering that $T - U$ is still a tensor (see below, the section on tensor algebra), we see that

if two tensors are equal in one frame, they are equal in any frame.

That is the basic value of tensors. They allow you to express equations of various kinds, equations of motion, equations of whatever it happens to be, in a form where the same exact equation will be true in any coordinate system. That is of course a deep advantage to thinking about tensors.

There are other objects which are not tensors. They will have the property that they may be zero in some frames and not zero in other frames. We are going to come across some of them.

Tensors have a certain invariance to them. Their components are not invariant. They change from one frame to another. But the statement that a tensor is equal to another tensor in frame independent.

Incidentally, when you write a tensor equation, the components have to match. It doesn't make sense to write an equation like W_q^p (where p is contravariant and q covariant) equals T^{pq} (where both indices are contravariant). Of course you can write whatever you like, but if, let's say in one coordinate system, the equation $W_q^p = T^{pq}$ happened to be true, then it would usually not be true in another. So normally we wouldn't write equations like that.

When thinking of two vectors, if we can write $V = W$, then they are equal in all coordinates systems. Note that in Euclidean geometry, or in non-Euclidean geometry with a positive definite distance, for $V = W$ to be true it is necessary and sufficient that the magnitude of $V - W$ be equal to zero. But this statement is not true in the Minkowski geometry of relativity, where the proper distance between two events may be zero without them being the same event.

In other words, notice that the magnitude of a vector and the vector itself are two different things. The magnitude of a vector is a scalar, whereas the vector is a complex object. It has components. It points in a direction. To say that two vectors are equal means that their magnitudes are the same and their directions are the same.

A tensor of higher rank is yet a more complicated object which points in several directions. It has got some aspect of it that points in one direction and some aspects that point in other directions. We are going to come to their geometry soon. But for the moment we define them by their transformation properties.

The next topic in tensor analysis is operations on tensors.

Tensor algebra

What can we do with tensors that make new tensors? We are not at this point interested in things that we can do to tensors which make other kinds of objects which are not tensors. We are interested in the operations we can do with tensors which will produce new tensors. In that way we can make a collection of things out of which we can build equations. And the equations will be the same in every reference frame.

First of all you can multiply a tensor by a numerical number. It is still a tensor. That rule is obvious and we don't need to spend time on it.

Then, we shall examine four operations. Most of them are very simple. The last one is not simple.

1. Addition of tensors. We can add two tensors of the same type, that is, of the same rank and the same numbers of contravariant and covariant indices. And addition of course includes also subtraction. If you multiply a tensor by a negative number and then add it, you are doing a subtraction.

2. Multiplication of tensors. We can multiply any pair

of tensors to make another tensor.

3. Contraction of a tensor. From certain tensors we can produce tensors of lower rank.

4. Differentiation of a tensor. But this will not be ordinary differentiation. It will be *covariant differentiation.* We will define it and see how it works.

Those are the four basic processes that you can apply to tensors to make new tensors. The first three are straightforward. As said, the last one is more intricate: differentiation with respect to what? Well, differentiation with respect to position. These tensors are things which might vary from place to place. They have a value at each point of the surface under consideration. They are tensor fields. At the next point on the surface they have a different value. Learning to differentiate them is going to be fun and hard. Not very hard, a little hard. Furthermore it belongs, strictly speaking, to tensor analysis and will be taken up in the next chapter.

Adding tensors: you only add tensors if their indices match and are of the same kind. For example if you have a tensor

$$T = T^{m...}_{...p}$$

with a bunch of upstairs contravariant indices, and a collection of downstairs covariant indices, and you have another tensor *of the same kind*

$$S = S^{m...}_{...p}$$

in other words their indices match exactly, then you are permitted to add them and construct a new sensor which we can denote

$$T + S$$

It is constructed in the obvious way: each component of the sum

$$(T + S)^{m...}_{...p}$$

is just the sum of the corresponding components of T and S. And it is obvious too to check that $T + S$ transforms as a tensor with the same rules as T and S. The same is true of $T - S$. It is a tensor. This is the basis for saying that tensor equations are the same in every reference frame – because $T - S = 0$ is a tensor equation.

Next, multiplication of tensors: now, unlike addition, multiplication of tensors can be done with tensors of any rank and type. The rank of a tensor is its number of indices. And we know that the two types, for each index, are contravariant or covariant. We can multiply $T^l_{\ mn}$ by $S^p_{\ q}$. The tensor multiplication being not much more than the multiplication of components and of the number of indices, we will get a tensor of the form P^{lp}_{mnq}.

Let's see a simple example: the tensor multiplication, also called *tensor product*, of two vectors. Suppose V^m is a vector with a contravariant index, and let's multiply it by a vector W_n with a covariant index. This produces a tensor with one upstairs index m and one downstairs index n

$$V^m \, W_n = T^m_n \tag{17}$$

A tensor is a set of values indexed by zero (in the case of a scalar), one (in the case of a vector) or several indices. This tensor T of equation (17) is a set of values – depending of the coordinate system in which we look at it – indexed by two indices m and n, respectively of contravariant and covariant type. It is tensor of rank two, contravariant for on index and covariant for the other.

We could have done the multiplication with some other vector X^n. And this would have produced some other tensor

$$V^m \ X^n = U^{mn} \tag{18}$$

We sometimes use the sign $\otimes$ to denote the tensor product. So equations (17) and (18) are sometimes written

$$V^m \otimes W_n = T^m_{\ n}$$

$$V^m \otimes X^n = U^{mn}$$

And this applies to the product of any tensors. The tensor product of two vectors is not their dot product. We will see how the dot product of two vectors is related to tensor algebra in a moment. With the tensor product we produce a tensor of higher rank, by just juxtaposing somehow all the components of the multiplicands.

How many components does $V^m \otimes X^n$ have? Since we are going to work mostly with 4-vectors in space-time, let's take V and X to be both 4-vectors. Each is a tensor of rank one with a contravariant index. Their tensor product U is a tensor of rank 2. It has 16 independent components, *each of them the simple multiplication of two numbers*

$$U^{11} = V^1 X^1, \ U^{12} = V^1 X^2, \ U^{13} = V^1 X^3, \ ...$$

$$... \ U^{43} = V^4 X^3, \ U^{44} = V^4 X^4$$

It is not the dot product. The dot product has only is one component, not sixteen. It is a number.

Sometimes the tensor product is called the outer product. But we shall continue to call it the tensor product of two tensors, and it makes another tensor.

Typically the tensor product of two tensors is a tensor of different rank than either one of the multiplicands.

The only way you can make a tensor of the same rank is for one of the factors to be a scalar. A scalar is a tensor of rank zero. You can always multiply a tensor by a scalar. Take any scalar S multiply it by, say, V^m. You get another tensor of rank one, i.e. another vector. It is simply V elongated by the value of S.

But generally you get back a tensor of higher rank with more indices obviously.

We are in the course of learning tensor algebra and tensor analysis. It is a bit dry. Where these tensors will come in? We will meet then in real life soon enough. But so far this is just a notational device.

Out of the four operations mentioned above, we already have addition and multiplication.

Let's now turn to contraction. Contraction is also an easy algebraic process. But in order to prove that the contraction of a tensor leads to a tensor we need a tiny little theorem. No mathematician would call it a theorem. They would at most call it maybe a lemma.

Here is what the lemma says. Consider the following quantity[2]

$$\frac{\partial X^b}{\partial Y^m} \frac{\partial Y^m}{\partial X^a} \tag{19}$$

[2]we begin to use also letters a, b, c, etc. for indices because there just aren't enough letters in the m range or the p range for our needs.

Remember that the presence of m upstairs and downstairs means implicitely that there is a sum to be perfomed over m. Expression (19) is the same as

$$\sum_m \frac{\partial X^b}{\partial Y^m} \frac{\partial Y^m}{\partial X^a} \tag{20}$$

What is the object in expression (19) or (20)? Do you recognize what it is? It is the change in X^b when we change Y^m a little bit, times the change in Y^m when you change X^a a little bit, summed over m. That is, we change Y^1 a little bit, then we change Y^2 a little bit, etc. What is expression (20) supposed to be?

Let's go over it in detail. Instead of X^b, consider any function F. Suppose F depends on $(Y^1,\ Y^2\ ...,\ Y^M)$, and each Y^m depends on X^a. Then, from elementary calculus,

$$\frac{\partial F}{\partial Y^m} \frac{\partial Y^m}{\partial X^a}$$

is nothing more than the partial derivative of F with respect to X^a (partial because there can be other X^n's on which the Y^m's depend). That is

$$\frac{\partial F}{\partial Y^m} \frac{\partial Y^m}{\partial X^a} = \frac{\partial F}{\partial X^a}$$

What if what F happens to be X^b? Well, there is nothing special in the formulas. We get

$$\frac{\partial X^b}{\partial Y^m} \frac{\partial Y^m}{\partial X^a} = \frac{\partial X^b}{\partial X^a}$$

But what is $\partial X^b / \partial X^a$? It looks like a stupid thing to look at. The X^n are independent variables, so the partial derivative of one with respect to another is either 1, if they are the same, that is if we are actually looking at $\partial X^a / \partial X^a$, or 0 otherwise. So $\partial X^b / \partial X^a$ is the Kronecker delta symbol. We shall denote it

$$\delta_a^b$$

Notice that we use an upper index and a lower index. We shall find out that δ_a^b itself happens to also be a tensor. That is a little weird because it is just a set of numbers. But it is a tensor with one contravariant and one covariant index.

Now that we have the little lemma we need in order to understand index contraction, let's do an example. And then define it more generally.

Let's take a tensor which is composed out of two vectors, one with a contravariant index and the other with a covariant index,

$$T_n^m = V^m \; W_n \tag{21}$$

Now what contraction means is: *take any upper index and any lower index and set them to be the same and sum over them.* In other words take

$$V^m \; W_m \tag{22}$$

This means $V^1 W_1 + V^2 W_2 + V^3 W_3 + ... + V^M W_M$, if M is the dimension of the space we are working with.

We have identified an upper index with a lower index. We are not allowed to do this with two upper indices. We are not allowed to do with two lower indices. But we can take an upper index and a lower index. And let's ask how expression (22) transforms.

Let's look at the transformation rule applied first to expression (21). We already know that it is a tensor. Here is how it transforms[3]

$$(V^m\ W_n)' = \frac{\partial Y^m}{\partial X^a}\ \frac{\partial X^b}{\partial Y^n}\ (V^a\ W_b) \tag{23}$$

Equation (23) is the transformation property of the tensor T^m_n which has one index upstairs and one index downstairs.

Now let $m = n$ and contract the indices. Remember, contracting means identifying an upper and a lower index *and sum over them.* So on the left hand side we get

$$(V^m\ W_m)'$$

How many indices does it have? Zero. The index m is summed over. The quantity is a scalar. It is by definition the expression of the scalar $V^m\ W_m$ in the prime coordinate system, which as we know doesn't change. So the contraction of $V^m\ W_n$ did create another tensor, namely a scalar.

We can check what equation (23) says. It should confirm that $(V^m\ W_m)'$ is the same as $V^m\ W_m$.

Now our little lemma comes in handy. On the right hand side of (23), when we set $m = n$ and sum over m, the sum of the products of partial derivatives is δ^b_a. So the right hand side is $V^a\ W_a$. But a or m are only dummy indices,

[3]We write $(V^m\ W_n)'$, but we could also write $(V^m)'\ (W_n)'$, because we know that they are the same. Indeed *that is what we mean when we say that the outer product of two vectors forms a tensor*: we mean that we can take the collection of products of their components in any coordinate system. Calculated in any two systems, $(V^m)'\ (W_n)'$ and $V^m\ W_n$ will be related by equation (23).

therefore equation (23) says indeed that

$$(V^m\ W_m)' = V^m\ W_m$$

So by contracting two indices of a tensor we make another tensor, in this case a scalar.

It is easy to prove, and the reader is encouraged to do it, that if you take any tensor with a bunch of indices, any number of indices upstairs and downstairs,

$$T^{nmr}_{pqs} \tag{24}$$

and you contract a pair of them (one contravariant and one covariant), say r and q, you get

$$T^{nmr}_{prs} \tag{25}$$

where the expression implicitely means a sum of components over r, and this is a new tensor.

Notice that the tensor of expression (24) has six indices, whereas the tensor of expression (25) has only four.

And notice two more things:

a) If we looked at $V^m\ W^n$, we would be dealing with a tensor which cannot be contracted. The analog of equation (23) would involve

$$\frac{\partial Y^m}{\partial X^a}\ \frac{\partial Y^n}{\partial X^b}$$

This quantity doesn't become the Kronecker delta symbol when we set $m = n$ and sum over it. And $\sum_m (V^m)'\ (W^m)'$ would not be equal to $\sum_m V^m\ W^m$.

b) The dot product of two vectors V and W is the contraction of the tensor $V^m\ W_n$. But in that case one vector must have a contravariant index, and the other a covariant index.

In other words, contraction is the generalization of the dot product (also called inner product) of two vectors. We are going to deal with inner products as soon as we work again with the metric tensor.

More on the metric tensor

The metric tensor plays a big role in Riemannian geometry. We showed its construction with the basis vectors e_m's, see figure 7 and after. In the set of equations (11), we wrote

$$g_{mn} = e_m.e_n$$

But let's now define it on its own terms abstractly. Again these are things we have already covered before, but let's do them again now we have a bit more practice with tensors.

The definition of the metric tensor goes like this. Consider a differential element dX^m which just represents the components of a displacement vector dX. In other words, we are at a point P on the Riemannian surface (or Riemannian space if we are in more that two dimensions), see figure 10, and we consider an infinitesimal displacement which we call dX – even though we also attach X to a specific coordinate system. We could call the small displacement dS but traditionally dS is a scalar representing a length.

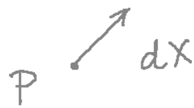

Figure 10: Displacement vector dX.

The contravariant components of dX are the coefficients of the vector dX in the expansion given by equation (4), which we rewrite below specifically for dX, supposing furthermore to make notations simple that there are three dimensions and therefore three axes,

$$dX = dX^1 e_1 + dX^2 e_2 + dX^3 e_3 \tag{26}$$

Each dX^m is a contravariant component of the little displacement vector of figure 10.

Now we ask: what is the length of that displacement vector?

Well, we need to know more about the geometry of the surface (also called *variety*) to know what the length of the little vector is. The surface or variety could be some arbitrarily shaped complicated space.

Specifying what the geometry of the variety is, in effect is specifying what the lengths of all the infinitesimal displacements are.

As said, we usually denote the length dS, and we usually work with its square. When the variety locally is Euclidean, dS is defined with Pythagoras theorem, but when the axes locally are not orthogonal or the dX^m are not expressed in units of length, or both, then Pythagoras theorem takes a more complicated form.

It is still quadratic in the dX^m's, but it may also involve products $dX^m dX^n$ and there is a coefficient g_{mn} in front of each quadratic term. The square of the length of any infinitesimal displacement is given by

$$dS^2 = g_{mn}\ dX^m dX^n$$

In general the g_{mn} depend on where we are, that is, they depend on P, which we locate with its coordinates X's. So we write more generally

$$dS^2 = g_{mn}(X)\ dX^m dX^n \tag{27}$$

We are going to stick with the case of four dimensions because we are in a course on relativity. For the moment, however, we don't consider the Einstein-Minkowski distance whose square can be a negative number. We are in a Riemannian geometry with four dimensions, where all distances are real and positive. In that case how many independent components are there in the g_{mn} object? Answer: to begin with there are 16, because g_{mn} is a 4 x 4 array.

But $dX^1 dX^2$ is exactly the same as $dX^2 dX^1$. So there is no point in having a separate variable for g_{12} and g_{21}, because they can be made equal to each other. So there are only 4 + 3 + 2 + 1 = 10 independent components in g_{mn} in four dimensions, see figure 11.

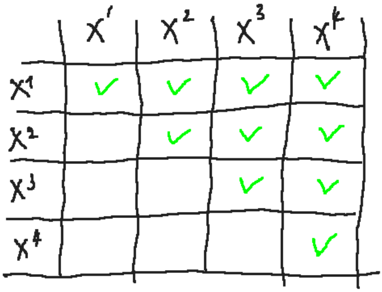

Figure 11: Independent components in g_{mn}.

Similarly in a three-dimensional space there would be 6 independent components in g_{mn}. And in two dimensions it would be 3.

So far we haven't proved that g_{mn} is a tensor. I called it the metric tensor, but let's now prove that it is indeed such an object. The basic guiding principle is that the length of a vector is a scalar, and that everybody agrees on that length. People using different coordinate systems won't agree on the components of the little vector dX (see figure 10), but they will agree on its length. Let's write again the length of dX, or rather its square

$$dS^2 = g_{mn}(X)\ dX^m\ dX^n \tag{28}$$

And now let's go from the X coordinates to the Y coordinates. Because dS^2 is invariant, the following holds

$$g_{mn}(X)\ dX^m\ dX^n = g'_{pq}(Y)\ dY^p\ dY^q \tag{29}$$

Now let's use this elementary calculus fact

$$dX^m = \frac{\partial X^m}{\partial Y^p}\ dY^p \tag{30}$$

And plug expression (30) for dX^m and for dX^n into equation (29). We get

$$g_{mn}(X)\,\frac{\partial X^m}{\partial Y^p}\,\frac{\partial X^n}{\partial Y^q}\,dY^p\,dY^q = g'_{pq}(Y)\,dY^p\,dY^q \tag{31}$$

The two sides of equation (31) are expressions of the same quadratic form in the dY^p's. That can only be true if the coefficients are the same. Therefore we established the following transformation property

$$g'_{pq}(Y) = g_{mn}(X)\,\frac{\partial X^m}{\partial Y^p}\,\frac{\partial X^n}{\partial Y^q} \tag{32}$$

This is just exactly the transformation property of a tensor with two covariant indices. So we discovered that the metric tensor is indeed really a tensor. It transforms as a tensor. This will have many applications.

The metric tensor has two lower indices because it multiplies the differential displacements dX^m's in equation (28) which have upper indices.

But the metric tensor is also just a matrix with $m\ n$ indices. Remembering that $g_{ij} = g_{ji}$, it is the following matrix, which we still denote g_{mn},

$$g_{mn} = \begin{pmatrix} g_{11} & g_{12} & g_{13} & g_{14} \\ g_{12} & g_{22} & g_{23} & g_{24} \\ g_{13} & g_{23} & g_{33} & g_{34} \\ g_{14} & g_{24} & g_{34} & g_{44} \end{pmatrix}$$

It is a symmetric matrix.

There is one more fact about this matrix, that is about the tensor g_{mn} thought of as a matrix. It has eigenvalues. And the eigenvalues are never zero.

The reason the eigenvalues are never zero is because a zero eigenvalue would correspond to a little eigenvector of zero length. But there are no vectors is 0 length. In Riemannian geometry every direction has a positive length associated with it.

What do we know about matrices which are symmetric and whose eigenvalues are all non-zero? Answer: they have inverses. The matrix of the metric tensor – both denoted g_{mn} or g for simplicity – has an inverse which in matrix algebra would be denoted g_{mn}^{-1} or simply g^{-1}. And

$$g^{-1}\ g = \text{the unit matrix}$$

In tensor algebra, the inverse matrix is not denoted g_{mn}^{-1} nor g^{-1}. It is denoted g^{mn}, with the indices upstairs.

g^{mn} is also a tensor. Its defining property is that, as a matrix, it is the inverse of the initial matrix g_{mn} with two lower indices. Let's write the corresponding equations. It is the last thing we shall do in this lesson. Let's do it slowly.

Consider two matrices A and B. Let's say two square matrices for simplicity, one denoted with upper indices and the other with lower indices

$$A^{mn} \text{ and } B_{pq}$$

How do we multiply them? That is, how do we compute the $m\ q$ component of the product? It is very simple. If you remember matrix algebra, it is

$$(AB)^m_q = A^{mr} B_{rq}$$

So let's compute the product of the two matrices g_{mn} by g^{mn}. By definition of g^{mn}, we have

$$g_{mn}\ g^{np} = \delta^p_m \tag{33}$$

where δ^p_m is the identity matrix.

Equation (33) is an equation in matrix algebra. But it is also an equation in tensor algebra. It is indeed elementary to show that g^{np} is also a tensor. Its expression in a Y coordinate system is by definition $(g')^{np}$, such that

$$g'_{mn}\ (g')^{np} = \delta^p_m$$

Then there are various mathematical ways to arrive at the analog of equation (32) for the tensor g with upper indices.

As a tensor equation, equation (33) shows on its left hand side the contraction of the tensor $g_{mn} \otimes g^{qp}$. And it says that the contraction of that product is the Kronecker delta object, which is necessarily also a tensor since it is the result of the contraction of a tensor.

g^{mn} is called the metric tensor with two contravariant indices.

The fact that there is a metric tensor with downstairs indices and a metric tensor with upstairs indices will play an important role.

So far everything we have seen on tensors was easy. It is essentially learning and getting accustomed with the notation.

Next step: differentiation of tensors

In the next chapter we will go on to the subject of curvature, parallel transport, differentiation of tensors, etc.

The idea of a covariant derivative will be a little more complicated than tensor algebra. Not much. But it is essential. We have to know how to differentiate things in a space, if we are going to do anything useful.

In particular, if we are going to study whether the space is flat, we have to know how things vary from point to point. The question of whether a space is flat or not fundamentally has to do with derivatives of the metric tensor – the character and nature of the derivatives of the metric tensor.

So in the next chapter we will talk a little bit about tensor calculus or tensor analysis, differentiation of tensors, and especially the notion of curvature.