Curvilinear coordinates, vectors, metric, connection

As we had discussed before, the a key idea in general relativity is to reproduce the effect of gravitation by curving the space-time manifold. To be able to understand the formulation of general relativity we must therefore learn to work with curved spaces. In this lecture we introduce the basic notions that characterise a curved manifold.

If we have a flat manifold then we can put Cartesian coordinates on the manifold: thus we would have x, y, z in flat 3-d space and x, y, z, t if we have both space and time in special relativity. The different coordinate directions are orthogonal to each other, and each direction extends to infinity on each side.

Even with a flat space, it is sometimes helpful to use coordinates that are not the Cartesian ones: an example is polar coordinates r, θ on a plane. Such coordinates are called *curvilinear* coordinates. If the space itself is curved, as for example the surface of a sphere, then we cannot put Cartesian coordinates at all. If the space has no particular symmetries, then no natural choice of coordinates will exist, and so we must study *general coordinate systems* on a manifold.

Recall that when we try to put coordinates θ, ϕ on the surface of a sphere, then these coordinates fail at the polar caps $\theta = 0, \pi$ (all values of ϕ represent the same point for these values of θ, a singular situation). Thus we will in general put coordinates on a 'patch' or small region of the space, and use a different patch to cover another region of of the space. These patches will overlap, so that no region of space is finally left uncovered.

Let the space be d dimensional. Some of these dimensions can be like time, or they can all be space directions; the difference will not be relevant at the moment. Thus our world is a 4-d manifold, with three space and one time direction at each point; these counts do not change when we consider curved spacetime instead of the flat one we considered in special relativity.

For a d-dimensional manifold we will introduce coordinates

$$\{x^1, x^2, \ldots x^d\} \tag{1}$$

Each point in the patch we are putting coordinates on will have a given value of these d numbers; conversly no two points in the patch will have the same value for all the coordinates. For convenience we often refer to the collection of all the coordinates of a point by just the symbol x: thus we would say 'the particle A is located at position x_1 and the particle B is located at x_2.

1. Functions

The simplest kind of quantity we can consider on this manifold is a *function* f. A real function f assigns a real number to each point of the patch, so that we will write

$$f = f(x^1, x^2, \dots x^d) \tag{1.1}$$

We may also consider complex functions, in which case the value of f at each point of the patch will be a complex number.

2. Contravariant vectors

The next object we consider is *vectors.* A vector V is based at a point x on the manifold. One may think of the vector as an arrow that leads from x to an infinitesimally separated point on the manifold; however we then also allow ourselves to 'multiply ' this arrow by an arbitrary number. Thus in fig. 2.1 the base of the arrow is at the point given by the coordinates

$$\{x^1, x^2, \dots, x^d\} \tag{2.1}$$

while the tip is at

$$\{x^1 + \delta x^1, x^2 + \delta x^2, \dots x^d + \delta x^d\} \tag{2.2}$$

Since the δx^i are infinitesimal, we multiply the above displacement by a large number $1/\epsilon$ (the quantity ϵ is also assumed infinitesimal). We now go through a proces of taking a limit: we keep the base of the arrow fixed, but let the tip be at

$$\{x^1 + \frac{\delta x^1}{2}, x^2 + \frac{\delta x^2}{2}, \dots x^d + \frac{\delta x^d}{2}\} \tag{2.3}$$

This time we multiply the arrow by $1/(\epsilon/2)$, thus obtaining 'approximately the same arrow. The limit of this process, as we take the arrow smaller and smaller, and the multiplying factor larger and larger, gives us the desired vector. Thus the vector is described by its components

$$V = \{V^1, V^2, \dots V^d\} \tag{2.4}$$

$$V^i = \lim_{\epsilon \to 0} \frac{\delta x^i}{\epsilon} \tag{2.5}$$

where the limit is understood in the sense described in words above.

We have now obtained a quantity, the vector, which is not the initial infinitesimal arrow. We cannot think of this vector anymore as leading from one point on the manifold to another point. Consider the example of a bead constrained to run along a curved wire (fig 2.1). If we want to find the velocity of the bead in the $x-y$ plane at a given point of time, we must take an arrow leading from the position of the bead at time t, to the point where the bead reaches at a time $t+\delta t$. We 'multiply' this arrow by $1/\delta t$, and take a limit where δt becomes infinitesimally small. This gives us the velocity vector, but this vector no longer leads from a point on the wire to another point on the wire. Instead it lives in the 'tangent space' to the wire at the point x where the velocity is measured. Similarily, we will have our vectors live in the tangent planes to the manifold; there is a d dimensional tangent plane at each point x of the manifold.

Let us now imagine a different set of coordinates covering the same patch, given by the numbers $\xi^i, i = 1 \ldots d$.

Since we already had the labels $x^i, i = 1 \ldots d$ describing each point of the patch we can regard each new coordinate as a function of the old coordinates

$$\xi^i = \xi^i(x^1, \ldots x^d), \quad i = 1 \ldots d \tag{2.6}$$

Thus we can define a matrix of partial derivatives

$$Q^i{}_j \equiv \frac{\partial \xi^i}{\partial x^j} \tag{2.7}$$

Each entry of this matrix is in general a nonconstant function over the patch.

By the chain rule for partial derivatives we have

$$\sum_{k=1}^{d} \frac{\partial \xi^i}{\partial x^k} \frac{\partial x^k}{\partial \xi^j} = \delta^i_j \tag{2.8}$$

Thus the partial derivatives $\frac{\partial x^i}{\partial \xi^j}$ define a matrix that is the inverse of the matrix Q.

We wish to ask the following question: given a vector with components V^i in the coordinate chart $\{x^i\}$, what will be the components of the vector in the chart $\{\xi^i\}$?

Using the notation used above in the description of a vector, let the base of the vector be at point x and the tip at point $x + \delta x$, in the unprimed coordinate system. In the primed coordinate system the base will be at $\xi(x)$, and the tip will be at $\xi(x + \delta x)$. We have

$$\xi^i(x + \delta x) = \xi^i + \sum_k \frac{\partial \xi^i}{\partial x^k}(x) \delta x^k \equiv \xi^i + \delta \xi^i \tag{2.9}$$

In the primed coordinate system the components of the vector V'^i will be given by the coordinate differences ξ^i between the tip and the base of the vector, together with the division by the infinitesimal parameter ϵ. Thus we get

$$V'^i = \frac{\delta\xi^i}{\epsilon} = \sum_k \frac{\partial\xi^i}{\partial x^k}(x)\frac{\delta x^k}{\epsilon} = \sum_k \frac{\partial\xi^i}{\partial x^k}(x)V^k \tag{2.10}$$

A collection of numbers labelled by one index (running over $1 \dots d$) which changes upon coordinate change in the manner given by (2.10) defines a *contravariant vector*. The prefix 'contravariant' is used to distinguish this quantity from another kind of vector that we will encounter, which we discuss next.

3. Covariant vectors

Let us now consider a function $f(x)$ and compute the *gradient* of f. This is the set of the following quantities

$$\{\frac{\partial f}{\partial x^1}, \frac{\partial f}{\partial x^2}, \dots \frac{\partial f}{\partial x^d}\} \tag{3.1}$$

We can denote the above array by the symbol W:

$$W = \{W_1, W_2, \dots W_d\} \tag{3.2}$$

The entries of the array are

$$W_i \equiv \frac{\partial f}{\partial x^i}, \quad i = 1 \dots d \tag{3.3}$$

Let us see what the gradiant looks like in the primed coordinate system. We have

$$W'_i \equiv \frac{\partial f}{\partial \xi^i} = \sum_k \frac{\partial f}{\partial x^k}\frac{\partial x^k}{\partial \xi^i} = \sum_k \frac{\partial x^k}{\partial \xi^i} W_k \tag{3.4}$$

While the similarity of (3.2) and (2.4) may make it appear that W is a vector just like V, we see from the transformation laws (3.4), (2.10) that this is not true! This may be confusing, for when we use Cartesian coordinates we are used to calling the gradient of a function a vector, just like any other vector. But the point here is that if we wish to remain within the class of Cartesian coordinate systems, then the maps from one such system to another is given by an *orthogonal* transformation:

$$QQ^T = Q^TQ = I \quad \rightarrow \quad \frac{\partial\xi^i}{\partial x^i} = \frac{\partial x^i}{\partial \xi^i} \tag{3.5}$$

and we find that the laws for transforming V and W become identical in this case.

4. Einstein summation convention

To distinguish contrvariant vectors from covariant vectors, we use the convention that the index for a contravariant vector is placed on top (e.g. V^i) while the index for a covariant vector is placed at the bottom (e.e. W_i).

We also observe that we have to perform at various points a summation $\sum_{k=1}^{d}$. But the index that is summed over appears exactly two times in the expression; in fact once as an 'up' index and once as a 'down' index. Further, we will find that whenever an index is repeated twice in an expression, i will always need to be summed over.

Given this observation about how indices appear when working with quantities in general coordinate systems, einstein introduced a convention: we do not write the summation sign explicitly, and assume that any index repeated twice is summed over (an index will never be repeated more than twice). If there is any exception to this rule that we wish to make when presenting an expression, then we manifestly say 'no summation over k'. The fact that indices always appear in this fashion has its roots in the geometric structure of objects described by curvilinear coordinates; this should become clearer as we encounter more examples.

5. The metric

A crucial element in describing the spacetime will be an expression for finding the distance between any two infinitesimally separated points on the manifold. A look at a general curved manifold soon indicates that it does not make sense to describe the distance between two arbitrary points in terms of their coordinates: in polar coordinates on the plane for example the point (r, θ) is the same as the point $(r, \theta+2\pi)$, so while the coordinate difference is nonzero, the actual separation between the points is zero. Instead, we will define the distance between infinitesimally separated points. The distance between two general points, if such a thing is to be defined, will have to include the specification of a path joining the points: along this path we would add up the infinitesimal lengths of the segments making up the path.

In 3-d Cartesian coordinates the distance ds between two infinitesimally separated points is given by

$$ds^2 = dx^2 + dy^2 + dz^2 \tag{5.1}$$

With polar coordinates on the plane the distance is

$$ds^2 = dr^2 + r^2 d\theta^2 \tag{5.2}$$

More generally, we will write

$$ds^2 = \sum_{i.j=1}^{d} g_{ij}(x) dx^i dx^j \ \rightarrow \ g_{ij} dx^i dx^j \tag{5.3}$$

where in the last step we have used the summation convention; we will not write the summation sign explicitly from now on.

It is clear that we can take the functions g_{ij} to be symmetric

$$g_{ij}(x) = g_{ji}(x) \tag{5.4}$$

since any antisymmetric part will cancel out in their definition (5.3). g_{ij} is called the *metric* on the space.

Let us see what the metric components will be in a new coordinate system. The distance ds between the two points will be the same, in any coordinate system. Thus we have

$$g_{ij} dx^i dx^j = g'_{kl} d\xi^k d\xi^l \tag{5.5}$$

But we can also write the LHS as

$$g_{ij} \frac{\partial x^i}{\partial \xi^k} d\xi^k \frac{\partial x^j}{\xi^l} d\xi^l \tag{5.6}$$

Then, comparing coefficients of $d\xi^k d\xi^l$ between (5.5) and (5.6) we get

$$g'_{kl} = g_{ij} \frac{\partial x^i}{\partial \xi^k} \frac{\partial x^j}{\xi^l} \tag{5.7}$$

We say that the metric tensor is a 'covariant tensor of rank 2', since when we change coordinates each index of the metric acts as if it was an index of a covariant vector. More generally, an object may have someindices up and some indices down; if it is a tensor then we can find its components in a transformed system of coordinates by treating each index as if it was a vector index of the appropriate type - contravariant if up and covariant if down. For example (x ate the unprimed coordinates and ξ are the primed coordinates)

$$T'^{i}_{jk} = \frac{\partial \xi^i}{\partial x^m} \frac{\partial x^n}{\partial \xi^j} \frac{\partial x^p}{\partial \xi^k} T^m_{np} \tag{5.8}$$

6. The inverse metric

We define an inverse g^{ij} to the metric g_{ij} through the equation

$$g^{ik} g_{kj} = \delta^i_j \tag{6.1}$$

We have written the indices of the inverse metric as if it was a contrvariant tensor of rank 2. Let us verify this. In a new coordinate system we will have

$$g'^{ik} g'_{kj} = \delta^i_j \tag{6.2}$$

Writing g'_{jk} in terms of the metric in the old coordinates we get

$$g'^{ik} \frac{\partial x^m}{\partial \xi^j} \frac{\partial x^n}{\partial \xi^k} g_{mn} = \delta^i_j \tag{6.3}$$

Let us 'strip off' one partial derivative from the LHS by multiplying both sides by $\frac{\partial \xi^j}{\partial x^l}$. The LHS gives

$$\frac{\partial \xi^j}{\partial x^l} g'^{ik} \frac{\partial x^m}{\partial \xi^j} \frac{\partial x^n}{\partial \xi^k} g_{mn} = \delta^m_l g'^{ik} \frac{\partial x^n}{\partial \xi^k} g_{mn} = g'^{ik} \frac{\partial x^n}{\partial \xi^k} g_{ln} \tag{6.4}$$

so that we get

$$g'^{ik} \frac{\partial x^n}{\partial \xi^k} g_{ln} = \frac{\partial \xi^j}{\partial x^l} \delta^i_j = \frac{\partial \xi^i}{\partial x^l} \tag{6.5}$$

Let us now multiply both sides by g^{lp}:

$$g'^{ik} \frac{\partial x^n}{\partial \xi^k} g_{ln} g^{lp} = g'^{ik} \frac{\partial x^n}{\partial \xi^k} \delta^p_n = g'^{ik} \frac{\partial x^p}{\partial \xi^k} = \frac{\partial \xi^i}{\partial x^l} g^{lp} \tag{6.6}$$

Let us now strip off the remaining partial derivative on the LHS by multiplying both sides by $\frac{\partial \xi^s}{\partial x^p}$:

$$g'^{ik} \frac{\partial x^p}{\partial \xi^k} \frac{\partial \xi^s}{\partial x^p} = g'^{ik} \delta^s_k = g'^{is} = \frac{\partial \xi^i}{\partial x^l} \frac{\partial \xi^s}{\partial x^p} g^{lp} \tag{6.7}$$

Thus we see that the inverse metric tensor indeed transforms as a contravariant tensor of rank 2.

7. The connection

Suppose we are trying to find the acceleration of a particle moving on a curved spacetime. At any given point on its path the particle will have a velocity vector that belongs to the tangent space at that point. At an infinitesimally later point on its path, the velocity vector will belong to the tangent space at the later position. We need to compare the latter velocity vector with the first one, in order to define the acceleration.

If we were in flat space with Cartesian coordinates there is no difficulty: we just look for the change in the components of the velocity vector, and do not worry about where each velocity vector is located. But such is not the case if we have curvilinear coordinates: the velocity vector at the two points might have different components because the velocity changed, or because the coordinate system used to define the components changed between the two points. For example consider the 2-dimensional plane and take two unit vectors pointing along the x direction: one at the point $(1,0)$ and the other at the point $(0,1)$, in Cartesian coordinates. Clearly these vectors are the 'same'. If we look at these vectors in polar coordinates, however, we find that the first has only a r component, while the second has only a θ component! Thus we cannot simply look at the components of a vector and decide if it is the same or not as a vector at a different point.

If we are in a space which is curved, then the situation is further complicated. We will make a rule to compare the vectors at infinitesimally separated points, so that we can say, given a vector at one point, what is the 'same' vector at the neighbouring point. But if we take a point that is a finite distance away, then we will have to 'transport' the initial vector along a succession of infinitesimal line segments to reach the final point. But the vector we obtain at this final point will in general depend on the path chosen between the points! Thus there is no no unique way to decide if vectors at two finitely separated points are the 'same'.

The above feature is not just a technical irritation in formulating vectors on curved space - it is the very basis of the definition of curvature. In fact a region of space is flat if and only if every vector at every point can be transported to any other point without the above ambiguity. Thus the rule for transport will be a very basic element in studying curved manifolds.

Let us see what form such a rule might take. Let V be a vector at a point x. We wish to ask: what is the 'same' vector at the point $x + \delta x$? We expect that the the vector at $x + \delta x$ will depend linearly on the vector at x; further, the change in the components between the two points will be proportional to the coordinate shifts δx^i. Thus we expect a rule like

$$V^i(x+dx) \equiv V^i(x) + \delta V^i \tag{7.1}$$

$$\delta V^i = -\Gamma^i_{jk}(x) V^j(x) \delta x^k \tag{7.2}$$

(We have put in the negative sign to accord with the definition of Γ that is standard.)

The set of functions Γ^i_{jk} are called the *connection*. The value of the connection is a part of the specification of the spacetime manifold. In a large class of physical settings, however, the connection is restricted to be a very special one: the Levi-civita connection, which we will now discuss.

Let us return for a moment to the case of curvilinear coordinates in flat space. Since the space is flat we can put Cartesian coordinates x^i on this space, and in these coordinates we transport a vector in the usual way: the components do not change under the transport. Thus the connection components Γ^i_{jk} all vanish. Now put coordinates ξ^i on the manifold. What will be the value of the ${\Gamma'}^i_{jk}$, the connection components in the new coordinates ξ^i?

Let us start with a vector with components V^i at the point x (x are Cartesian coordinates). At point $x + \delta x$ the transported vector also has components V^i. At the point x the vector components in the new ξ coordinate system are

$$V'^i(x) = \frac{\partial \xi^i}{\partial x^m}(x) V^m \tag{7.3}$$

At the point $x + \delta x$ the components are

$$V'^i(x+\delta x) = \frac{\partial \xi^i}{\partial x^m}(x+\delta x) V^m = \frac{\partial \xi^i}{\partial x^m}(x) V^m + \frac{\partial^2 \xi^i}{\partial x^m \partial x^l} V^m \delta x^l \tag{7.4}$$

Thus the change of components is

$$\delta V'^i = \frac{\partial^2 \xi^i}{\partial x^m \partial x^l} V^m \delta x^l = \frac{\partial^2 \xi^i}{\partial x^m \partial x^l} \frac{\partial x^m}{\partial \xi^j} V'^j \delta x^l = \frac{\partial^2 \xi^i}{\partial x^m \partial x^l} \frac{\partial x^m}{\partial \xi^j} \frac{\partial x^l}{\partial \xi^k} V'^j \delta \xi^k \tag{7.5}$$

Thus the connection is

$$\Gamma'^i_{jk} = -\frac{\partial^2 \xi^i}{\partial x^m \partial x^l} \frac{\partial x^m}{\partial \xi^j} \frac{\partial x^l}{\partial \xi^k} \tag{7.6}$$

To make sense of this expression we would like to relate it to expressions for other geometric quantities on the manifold. Let us assume that the Cartesian space we started with was the usual kind of space - namely it had no 'time' coordinartes. Then the metric in cartesian coordinates is just $g_{ij} = \delta_{ij}$. In the ccordinates ξ the metric will then be

$$g'_{pq} = \frac{\partial x^r}{\partial \xi^p}\frac{\partial x^s}{\partial \xi^q} g_{rs} = \frac{\partial x^r}{\partial \xi^p}\frac{\partial x^s}{\partial \xi^q}\delta_{rs} = \frac{\partial x^r}{\partial \xi^p}\frac{\partial x_r}{\partial \xi^q} \tag{7.7}$$

We can make second derivatives from this expression by differentiation

$$\frac{\partial g'_{pq}}{\partial \xi^s} = \frac{\partial^2 x^r}{\partial \xi^p \partial \xi^s}\frac{\partial x_r}{\partial \xi^q} + \frac{\partial x^r}{\partial \xi^p}\frac{\partial^2 x_r}{\partial \xi^q \xi^s} \tag{7.8}$$

However these are the 'wrong' partial derivatives; in (7.6) we need second partial derivatives of the ξ coordinates, not the x coordinates. But we have from the chain rule for partial derivatives:

$$\frac{\partial \xi^p}{\partial x^r}\frac{\partial x^r}{\partial \xi^q} = \delta^p_q \tag{7.9}$$

and by differentiation with respect to x^s we get

$$\frac{\partial^2 \xi^p}{\partial x^r \partial x^s}\frac{\partial x^r}{\partial \xi^q} + \frac{\partial \xi^p}{\partial x^r}\frac{\partial^2 x^r}{\partial \xi^q \partial \xi^t}\frac{\partial \xi^t}{\partial x^s} = 0 \tag{7.10}$$

(We had to use the chain rule for differentiation in the second term.) Using the above relation we find

$$\Gamma'^i_{jk} = \frac{\partial \xi^i}{\partial x^m}\frac{\partial^2 x^m}{\partial \xi^j \partial \xi^t}\frac{\partial \xi^t}{\partial x^l}\frac{\partial x^l}{\partial \xi^k} = \frac{\partial \xi^i}{\partial x^m}\frac{\partial^2 x^m}{\partial \xi^j \partial \xi^t}\delta^t_k = \frac{\partial \xi^i}{\partial x^m}\frac{\partial^2 x^m}{\partial \xi^j \partial \xi^k} \tag{7.11}$$

We now have the same kind of second order partial derivative as the one occuring in (7.8), but the latter contains a sum of two such partial derivatives. To get rid of the last term on the RHS of (7.8) we can antisymmetrise

$$\frac{\partial g'_{pq}}{\partial \xi^s} - \frac{\partial g'_{ps}}{\partial \xi^q} = \frac{\partial^2 x^r}{\partial \xi^p \partial \xi^s}\frac{\partial x_r}{\partial \xi^q} - \frac{\partial^2 x^r}{\partial \xi^p \partial \xi^q}\frac{\partial x_r}{\partial \xi^s} \tag{7.12}$$

We still have two terms with second order partial derivatives, but we can consider

$$\frac{\partial g'_{pq}}{\partial \xi^s} - \frac{\partial g'_{ps}}{\partial \xi^q} + \frac{\partial g'_{qs}}{\partial \xi^p} = 2\frac{\partial^2 x^r}{\partial \xi^p \partial \xi^s}\frac{\partial x_r}{\partial \xi^q} \tag{7.13}$$

To relate this expression to (7.11) we remove the first derivative term on the RHS of the above equation:

$$[\frac{\partial g'_{pq}}{\partial \xi^s} - \frac{\partial g'_{ps}}{\partial \xi^q} + \frac{\partial g'_{qs}}{\partial \xi^p}]\frac{\partial \xi^q}{\partial x^t} = 2\frac{\partial^2 x^r}{\partial \xi^p \partial \xi^s}\frac{\partial x_r}{\partial \xi^q}\frac{\partial \xi^q}{\partial x^t} = 2\frac{\partial^2 x^r}{\partial \xi^p \partial \xi^s}\delta^r_t = 2\frac{\partial^2 x^t}{\partial \xi^p \partial \xi^s} \tag{7.14}$$

and then replace it with a different first derivative term

$$[\frac{\partial g'_{pq}}{\partial \xi^s} - \frac{\partial g'_{ps}}{\partial \xi^q} + \frac{\partial g'_{qs}}{\partial \xi^p}]\frac{\partial \xi^q}{\partial x^t}\frac{\partial \xi^u}{x^t} = 2\frac{\partial^2 x^t}{\partial \xi^p \partial \xi^s}\frac{\partial \xi^u}{x^t} \tag{7.15}$$

But recall that since the unprimed coordiates are Cartesian, $g^{ij} = \delta^{ij}$. So

$$\frac{\partial \xi^q}{\partial x^t}\frac{\partial \xi^u}{x^t} = g'^{qu} \tag{7.16}$$

Thus we get

$$[\frac{\partial g'_{pq}}{\partial \xi^s} - \frac{\partial g'_{ps}}{\partial \xi^q} + \frac{\partial g'_{qs}}{\partial \xi^p}]g'^{qu} = 2\frac{\partial^2 x^t}{\partial \xi^p \partial \xi^s}\frac{\partial \xi^u}{x^t} \tag{7.17}$$

Substituting this in (7.11) we finally get

$$\Gamma'^i_{jk} = \frac{1}{2}g'^{il}[g'_{lj,k} + g'_{lk,j} - g'_{jk,l}] \tag{7.18}$$

The above expression was derived for the connection in curvilinear coordinates on flat space. but since this expression involves the metric, which is a geometric quantity on general curved pacetime, we postulate that the same expression for the connection in terms of the metric will hold in curved space as well. To see if the postulate makes sense, we will examine various properties of the connection defined by such a relation, and judge whether they are appropriate for the connection we desire.