6 Covariant differentiation

Any dynamical physical theory must deal in time varying quantities, and if this theory is also to be relativistic, spatially varying quantities too. Since GR is a covariant theory, we are confronted with the problem of constructing quantities that represent rates of change, but which can be defined in any coordinate system. In other words, we need to define a *derivative* which transforms covariantly under a general coordinate transformation: we call this the **covariant derivative**.

For any scalar function, say ϕ, defined on the manifold, the partial derivative

$$\phi_{,\nu} \equiv \frac{\partial \phi}{\partial x^{\nu}}$$

transforms as a $(0,1)$ tensor, i.e.

$$\phi'_{,\nu} \equiv \frac{\partial \phi'(x')}{\partial x'^{\nu}} = \frac{\partial x^{\mu}}{\partial x'^{\nu}}\frac{\partial \phi(x)}{\partial x^{\mu}} \equiv \frac{\partial x^{\mu}}{\partial x'^{\nu}}\phi_{,\mu} \tag{128}$$

Mathematical Aside: the notation in equation (128) may appear a little awkward. A scalar function simply maps each point, P, of the manifold to a real number. If the coordinates of P are $\{x^1, x^2, ..., x^n\}$ in an unprimed coordinate system and $\{x'^1, x'^2, ..., x'^n\}$ in another, primed, coordinate system, then the *form* of the scalar function may be different in the two coordinate systems, although $\phi(x^1, x^2, ..., x^n) = \phi'(x'^1, x'^2, ..., x'^n)$. (We saw this in equation (87) of Section 4 above). Consider, for example, $\phi(x, y) = x^2 + y^2$, defined on $\mathbf{E^2}$. Then $\phi(x, y) = \phi'(r, \theta) = r^2$.

If we take the second derivatives of ϕ, however, the quantities $\phi_{\ ,\nu\mu}$ do *not* have the transformation properties of a tensor. Indeed in general if T^{ijk}_{lmn} is a tensor, $T^{ijk}_{lmn,p}$ will usually not be. Consider for instance $A^i_{\ ,j}$.

$$\begin{aligned}
A'^i_{\ ,j} &= \frac{\partial A'^i(x')}{\partial x'^j} \\
&= \frac{\partial}{\partial x'^j}\left(\frac{\partial x'^i}{\partial x^k}A^k\right) \\
&= \frac{\partial x^l}{\partial x'^j}\frac{\partial}{\partial x^l}\left(\frac{\partial x'^i}{\partial x^k}A^k\right) \\
&= \frac{\partial x'^i}{\partial x^k}\frac{\partial x^l}{\partial x'^j}A^k_{\ ,l} + \frac{\partial x^l}{\partial x'^j}\frac{\partial^2 x'^i}{\partial x^l \partial x^k}A^k
\end{aligned} \tag{129}$$

Evidently the presence of the second term of equation (129),

$$\frac{\partial x^l}{\partial x'^j}\frac{\partial^2 x'^i}{\partial x^l \partial x^k}A^k$$

is the reason why $A^i_{,j}$ does not transform as a tensor. Under linear transformations $A^i_{,j}$ will behave like a tensor, since for linear transformations

$$\frac{\partial^2 x'^i}{\partial x^l \partial x^k} = 0$$

However, under nonlinear transformations this will not be the case.

The question is how can we design a derivative-like quantity that *does* transform as a tensor?

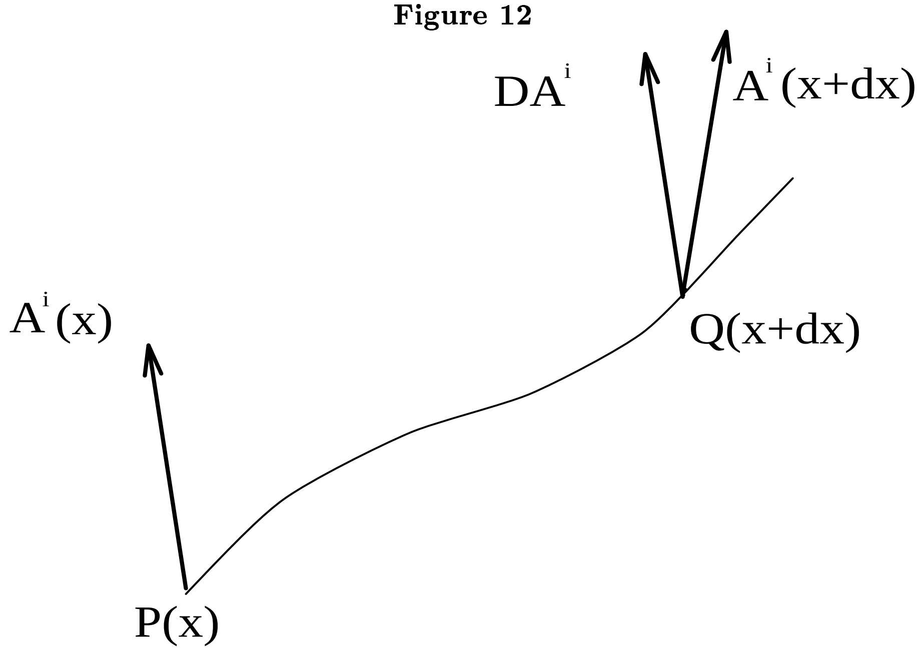
6.1 Parallel transport

The root of the problem is that in obtaining the derivative of A^i with respect to some coordinate, x, say, we compute the difference between the components $A^i(x)$ at P, and $A^i(x + dx)$ at neighbouring point Q.

> **Mathematical note:** we have simplified our notation here a little for clarity. Suppose the coordinate with respect to which we are differentiating is $x = x^1$. In fact we compute the difference between the components $A^i(x^1, x^2, ..., x^n)$ at P and $A^i(x^1 + dx^1, x^2, ..., x^n)$ at Q – i.e. with all other coordinates held fixed. For simplicity we write these expressions simply as $A^i(x)$ and $A^i(x + dx)$ respectively.

Now this difference, $A^i(x+dx) - A^i(x)$, is **not** a vector (i.e. a $(1, 0)$ tensor), since $A^i(x+dx)$ transforms as a vector component at Q and $A^i(x)$ as a vector component at P. To overcome this problem we need to introduce the important notion of **parallel transport**.

Figure 12



The idea (see Figure 12) is to displace or transport the vector from P to Q. We require this displaced quantity to transform as a vector at Q. How can this be done? We wish to construct a vector, $\vec{DA}$, at Q which is somehow related to $\vec{A}$ at P. We shall think of this as a **transport**, or displacement, of $\vec{A}$ from P to Q. We write the displaced vector $\vec{DA}$ at Q in component form as $DA^i(x+dx)$. From the linearity of tensors the difference, $A^i(x+dx) - DA^i(x+dx)$ will then transform as a vector at Q.

For this displacement to fit in with the notion of 'parallel' which we carry over from our everyday experience of flat space we should like the process or mapping to be *linear*. Hence, if $\vec{A}$ and $\vec{B}$ are transported into $\vec{DA}$ and $\vec{DB}$ respectively, then $a\vec{A}+b\vec{B}$ should be transported into $a\vec{DA}+b\vec{DB}$. We would also like the mapping to be unique – i.e. no two different vectors at P are mapped into the same vector at Q. Later, when we consider the particular case of Riemannian manifolds, we shall impose some further conditions on this transport – e.g. that it preserves the magnitude of vectors and angles between vectors – but for the time being we only demand that it is linear and unique.

If we write

$$DA^i(x+dx) = A^i(x) + \delta A^i(x) \tag{130}$$

then linearity requires $\delta A^i(x)$ to be linearly dependent on the components A^i. We should also expect it to depend linearly on the coordinate displacement, dx^k, between P and Q. From these considerations, we can in fact guess the form of $\delta A^i(x)$ to be

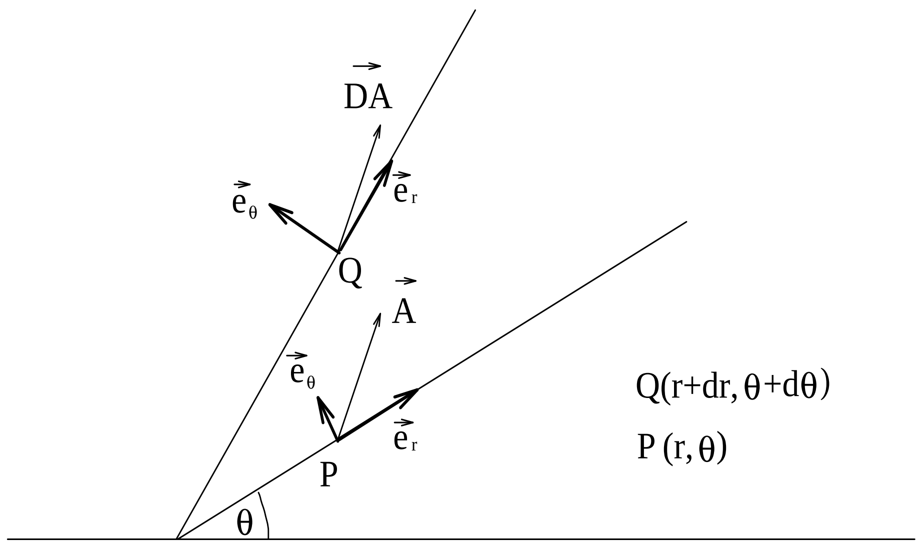
$$\delta A^i(x) = -\Gamma^i_{jk} A^j dx^k \tag{131}$$

(The minus sign here is just to accord with future definition of what are called the *affine connections* or *Christoffel symbols*. Usually the term Christoffel symbol is reserved for the particular case of a Riemannian space, where – as we shall see – they can be expressed in terms of the metric tensor).

6.2 Parallel transport in $\mathbf{E}^2$

To justify the form of equation (131) consider parallel transport in $\mathbf{E^2}$. In $\mathbf{E^2}$ the displaced vector $\vec{DA}$ is, in fact, *equal* to $\vec{A}$ (see Figure 13).

Figure 13



However, even in $\mathbf{E^2}$, the components of $\vec{DA}$ at Q, in an arbitrary coordinate system, will not necessarily be the same as those of $\vec{A}$ at P. In a Cartesian coordinate system we expect that $DA^i = A^i$, but this will not be true for all coordinate systems. Consider, for example, polar coordinates $\{r, \theta\}$. We can define basis vectors $\{\vec{e}_\theta, \vec{e}_r\}$ by

$$\vec{e}_\theta = \frac{\partial \vec{r}}{\partial \theta}, \quad \vec{e}_r = \frac{\partial \vec{r}}{\partial r} \tag{132}$$

where

$$\vec{r} = r\cos\theta\,\vec{i} + r\sin\theta\,\vec{j} \tag{133}$$

and $\{\vec{i}, \vec{j}\}$ are the Cartesian basis vectors for $\mathbf{E^2}$. It is then easy to show that

$$\vec{e}_\theta = -r\sin\theta\,\vec{i} + r\cos\theta\,\vec{j} \tag{134}$$

$$\vec{e}_r = \cos\theta\,\vec{i} + \sin\theta\,\vec{j} \tag{135}$$

Now since $\vec{DA} = \vec{A}$, we must have

$$(DA^i)\,\vec{e}_i(x+dx) \;=\; A^i\,\vec{e}_i(x) \tag{136}$$

where $\{\vec{e}_i(x)\}$ are basis vectors at P and $\{\vec{e}_i(x+dx)\}$ are basis vectors at Q (in general different from those at P). Writing

$$\vec{e}_i(x+dx) \;=\; \vec{e}_i(x) \;+\; \frac{\partial \vec{e}_i}{\partial x^k}dx^k \tag{137}$$

and

$$DA^i \;=\; A^i(x) + \delta A^i(x) \tag{138}$$

and substituting into equation (136) yields, to first order,

$$\delta A^i(x)\ \vec{e}_i\ +\ \frac{\partial \vec{e}_i}{\partial x^k} A^i dx^k\ =\ 0 \tag{139}$$

Evidently the vector

$$\frac{\partial \vec{e}_i}{\partial x^k}$$

can itself be expressed in terms of the basis vectors at P, and we write (with foresight)

$$\frac{\partial \vec{e}_i}{\partial x^k}\ =\ \Gamma^j_{ik}\ \vec{e}_j \tag{140}$$

Substituting equation (140) into equation (137) we obtain

$$(\delta A^j(x) + \Gamma^j_{ik} A^i dx^k)\ \vec{e}_j\ =\ 0 \tag{141}$$

This is a vector equation, and necessarily each component must be zero. Thus we obtain

$$\delta A^j(x) = -\Gamma^j_{ik}\ A^i dx^k \tag{142}$$

Reorganising the indices we can write this precisely in the form of equation (131).

Consider, as an example, the case of polar coordinates. We have

$$\frac{\partial \vec{e}_\theta}{\partial \theta}\ =\ -r\cos\theta\ \vec{\imath} - r\sin\theta\ \vec{\jmath} = -r\vec{e}_r \Rightarrow \Gamma^\theta_{\theta\theta} = 0, \quad \Gamma^r_{\theta\theta} = -r \tag{143}$$

$$\frac{\partial \vec{e}_\theta}{\partial r}\ =\ -\sin\theta\ \vec{\imath} + \cos\theta\ \vec{\jmath} = \frac{\vec{e}_\theta}{r} \Rightarrow \Gamma^\theta_{\theta r} = \frac{1}{r}, \quad \Gamma^r_{\theta r} = 0 \tag{144}$$

$$\frac{\partial \vec{e}_r}{\partial \theta}\ =\ -\sin\theta\ \vec{\imath} + \cos\theta\ \vec{\jmath} = \frac{\vec{e}_\theta}{r} \Rightarrow \Gamma^\theta_{r\theta} = \frac{1}{r}, \quad \Gamma^r_{r\theta} = 0 \tag{145}$$

$$\frac{\partial \vec{e}_r}{\partial r}\ =\ 0 \Rightarrow \Gamma^\theta_{rr} = 0, \quad \Gamma^r_{rr} = 0 \tag{146}$$

Notice that here $\Gamma^i_{jk} = \Gamma^i_{kj}$, i.e. the Christoffel symbols are symmetric in their lower indices.

Of course the Christoffel symbols, Γ^i_{jk}, will be different when evaluated in different coordinate systems. In the above example we evaluated the Christoffel symbols in polar coordinates; for Cartesian coordinates, on the other hand, they will all be zero. However we want the components DA^i to transform as vector components at Q. This obviously imposes certain constraints on the transformation properties of the Christoffel symbols. In particular Γ^i_{jk} do **not** transform as tensor components (see Tutorial Sheet 4). The transformation law for the Christoffel symbols is given by

$$\Gamma'^i_{jk} = \frac{\partial x'^i}{\partial x^r}\frac{\partial x^s}{\partial x'^j}\frac{\partial x^t}{\partial x'^k}\Gamma^r_{st} + \frac{\partial x'^i}{\partial x^l}\frac{\partial^2 x^l}{\partial x'^j \partial x'^k} \tag{147}$$

6.3 Parallel transport for general manifolds

The preceding section showed how we may define parallel transport from an arbitrary point P to neighbouring point Q in $\mathbf{E^2}$. In fact we defined parallel transport in the obvious way, by requiring that the parallel transported vector, $\vec{D}A$, at Q is identically equal to the original vector, $\vec{A}$, at P. Equation (139) then expressed the difference between the components, A^i, at P and the parallel transported components, DA^i, at Q, with respect to an arbitrary basis, in terms of the **connection symbols**, Γ^i_{jk}.

As a consequence of how we defined parallel transport, the connection symbols ensure that the DA^i transform as contravariant components at Q when the A^i transform as contravariant components at P. Connection symbols which have this property define what is known as an **affine connection**, and a general manifold on which we have defined an affine connection is called an **affine manifold**.

6.4 Parallel transport of a scalar

We can also define the parallel transport of a scalar from P, with coordinates x, to Q, with coordinates $x + dx$. Thus if ϕ is a scalar function on the manifold, we define

$$D\phi(x + dx) = \phi(x) \tag{148}$$

With this trivial definition we have

$$\phi(x + dx) - D\phi(x + dx) = \phi(x + dx) - \phi(x) \simeq \frac{\partial\phi}{\partial x^k} dx^k \tag{149}$$

This quantity is evidently a scalar. No matter what coordinate system it is evaluated in it is always the same – it is simply the difference between the value of the scalar function at P and at Q. It follows that the quantity $\frac{\partial\phi}{\partial x^k}$ must therefore transform as a $(0, 1)$ tensor, since dx^k transforms as a $(1, 0)$ tensor and the LHS of equation (149) is an invariant. (In fact we already saw that $\frac{\partial\phi}{\partial x^k}$ transforms as a $(0, 1)$ tensor in equation (128) above, from the chain rule for partial differentiation).

6.5 Parallel transport of a tensor

The notion of parallel transport can easily be extended to tensors of arbitrary rank, although the algebra becomes somewhat messier for higher rank tensors. However, once we have defined parallel transport of a contravariant vector through equations (138) to (142), and a scalar through equation (148), parallel transport of an arbitrary tensor can only take one form.

6.5.1 Parallel transport of $(0,1)$ tensor

To see how this generalisation works let us first consider the covariant components, B_i, of a $(0,1)$ tensor – i.e. a one-form. In a similar manner to our treatment of vectors, let us write the parallel transport of an arbitrary one form in terms of its covariant components B_i, viz.

$$DB_i(x+dx) = B_i(x) + \delta B_i(x) \tag{150}$$

Now since B_iA^i is a scalar, for arbitrary A^i, it must be displaced as a scalar. Thus, from equation (148)

$$D(B_iA^i) = (B_iA^i) + \delta(B_iA^i) = B_iA^i \tag{151}$$

Thus

$$\delta(B_iA^i) = 0 \tag{152}$$

from which it follows that

$$B_i\delta A^i + \delta B_iA^i = 0 \tag{153}$$

From equations (142) and (153) we obtain

$$\delta B_i = \Gamma^j_{ik}B_jdx^k \tag{154}$$

from which an expression for $DB_i(x+dx)$ then follows.

6.5.2 Parallel transport of a $(2,0)$ tensor

Consider now the parallel transport of a $(2,0)$ tensor, whose components are T^{ij} in some coordinate frame. In a similar manner to our treatment of vectors and one-forms, we write

$$DT^{ij}(x+dx) = T^{ij}(x) + \delta T^{ij}(x) \tag{155}$$

Now take two arbitrary $(0,1)$ tensors with (covariant) components B_i and C_i. Since $B_iC_jT^{ij}$ is a scalar, we must have

$$D(B_iC_jT^{ij}) = B_iC_jT^{ij} + \delta(B_iC_jT^{ij}) = B_iC_jT^{ij} \tag{156}$$

Thus it follows that

$$B_iC_j\delta T^{ij} + B_i\delta C_jT^{ij} + \delta B_iC_jT^{ij} = 0 \tag{157}$$

Now from equation (154),

$$\delta B_i = \Gamma^n_{ik}B_ndx^k$$

and

$$\delta C_j = \Gamma^m_{jk}C_mdx^k$$

Substituting into equation (157) we obtain

$$B_i C_j \delta T^{ij} + B_i \Gamma^m_{jk} C_m dx^k T^{ij} + \Gamma^n_{ik} B_n dx^k C_j T^{ij} = 0 \tag{158}$$

Rearranging and changing the dummy indices we get

$$B_i C_j (\delta T^{ij} + \Gamma^j_{mk} T^{im} dx^k + \Gamma^i_{nk} T^{nj} dx^k) = 0 \tag{159}$$

Since B_i and C_j were arbitrary $(0,1)$ tensors, the term in the bracket must be zero. Thus

$$\delta T^{ij} = -\Gamma^j_{mk} T^{im} dx^k - \Gamma^i_{nk} T^{nj} dx^k \tag{160}$$

Exactly the same reasoning can be applied to a tensor of arbitrary rank. Thus for the tensor D^{ij}_{kl} we have

$$\delta D^{ij}_{kl} = -\Gamma^i_{mp} D^{mj}_{kl} dx^p \; - \Gamma^j_{mp} D^{im}_{kl} dx^p \; + \Gamma^m_{kp} D^{ij}_{ml} dx^p \; + \Gamma^m_{lp} D^{ij}_{km} dx^p \tag{161}$$

Exercise: Write down an expression for δP^{klm}_i

6.6 Covariant derivative

Having defined the parallel transport of scalars, vectors one-forms and general tensors, we now have a means to define a derivative-like quantity which transforms like a tensor – i.e. the **covariant derivative**.

6.6.1 Covariant differentiation of a scalar

We have already noted above in equation (128) that $\partial\phi/\partial x^k$ transforms as a $(0,1)$ tensor. (We also obtained the same result in equation (149), after applying a parallel displacement to ϕ). Thus we define the covariant derivative of ϕ simply to be equal to the partial derivative of ϕ.

6.6.2 Covariant differentiation of a one-form

Consider first a $(0,1)$ tensor, $\mathbf{B}$ defined over the manifold. In some coordinate system and coordinate basis the components may be written B_i. Consider two points P and Q with coordinates x^k and $x^k + dx^k$ respectively. If we have defined parallel transport in terms of an affine connection, then the quantity $B_i(x+dx) - DB_i(x+dx)$ also transforms as a $(0,1)$ tensor. Substituting from equation (150) we obtain

$$B_i(x+dx) - DB_i(x+dx) = B_i(x+dx) - B_i(x) - \delta B_i(x) \tag{162}$$

We can rewrite the first two terms of the right hand side as

$$B_i(x + dx) - B_i(x) = \frac{\partial B_i}{\partial x^k} dx^k = B_{i,k} dx^k \tag{163}$$

Thus

$$B_i(x + dx) - DB_i(x + dx) = (B_{i,k} - \Gamma^j_{ik} B_j) dx^k \tag{164}$$

This implies that $B_{i,k} - \Gamma^j_{ik} B_j$ transforms as a $(0, 2)$ tensor, since the LHS is a $(0, 1)$ tensor and dx^k is $(1, 0)$ tensor. We call $B_{i,k} - \Gamma^j_{ik} B_j$ the **covariant derivative** of the B_i, and denote covariant differentiation by a semi-colon, i.e.

$$B_{i;k} = B_{i,k} - \Gamma^j_{ik} B_j \tag{165}$$

As an example of covariant differentiation, consider the $(0, 1)$ tensor

$$\frac{\partial \phi}{\partial x^i} = \phi_{\ ,i}$$

We can now define the covariant derivative of $\phi_{\ ,i}$ to be

$$\phi_{,i;k} = \phi_{,ik} - \Gamma^j_{ik} \phi_{,j} \tag{166}$$

It is easy to show (see Tutorial Sheet 4) that $\phi_{,ik} - \phi_{,ki}$ transforms as a $(0, 2)$ tensor. If we choose the affine connection to be symmetric – i.e. $\Gamma^j_{ik} = \Gamma^j_{ki}$ – then

$$\phi_{,i;k} - \phi_{,k\ ;i} = \phi_{,ik} - \phi_{,ki} \tag{167}$$

A space for which this is the case is called **torsion free**.

6.6.3 Covariant differentiation of a vector

For a $(1, 0)$ tensor, the covariant derivative can be defined in an analogous way to that for a one-form. Suppose that a vector field $\vec{A}$ is defined on the manifold. Choosing coordinates and the corresponding coordinate basis, we have

$$A^i(x + dx) - DA^i(x + dx) = A^i + A^i_{,k} dx^k - (A^i - \Gamma^i_{jk} A^j dx^k) = (A^i_{,k} + \Gamma^i_{jk} A^j) dx^k \tag{168}$$

The LHS of equation (168) transforms as a contravariant vector, as does dx^k. Thus $A^i_{,k} + \Gamma^i_{jk} A^j$ must transform as a $(1, 1)$ tensor.

We write

$$A^i_{;k} = A^i_{,k} + \Gamma^i_{jk} A^j \tag{169}$$

and refer to $A^i_{;k}$ as the covariant derivative of A^i with respect to x^k.

6.6.4 Covariant differentiation of tensor

Since we have defined parallel transport for a tensor of arbitrary rank, it is straightforward to define covariant differentiation of such a tensor. For example, the covariant derivative of a $(3, 3)$ tensor is given by

$$T^{ijk}_{lmn;p}dx^p = T^{ijk}_{lmn}(x + dx) - DT^{ijk}_{lmn}(x + dx) = T^{ijk}_{lmn,p}dx^p - \delta T^{ijk}_{lmn} \tag{170}$$

The last term on the right can easily (if lengthily!) be expressed in terms of the affine connections, so that equation (170) can be reduced to

$$T^{ijk}_{lmn;p} = T^{ijk}_{lmn,p} + \Gamma^i_{rp}T^{rjk}_{lmn} + \Gamma^j_{rp}T^{irk}_{lmn} + \Gamma^k_{rp}T^{ijr}_{lmn} - \Gamma^r_{lp}T^{ijk}_{rmn} - \Gamma^r_{mp}T^{ijk}_{lrn} - \Gamma^r_{np}T^{ijk}_{lmr} \tag{171}$$

6.7 Christoffel symbols

So far we have defined the Christoffel symbols without any reference to the metric tensor, g_{ij}. We simply assumed that some parallel transport of vectors and scalars was defined on the manifold, and showed that this notion of parallel transport could naturally be extended to tensors of arbitrary rank. However, if a metric is defined on the manifold there is a very natural definition of the Christoffel symbols, or affine connections, in terms of g_{ij}. The definition is as follows.

Suppose we require that the magnitude of a parallel displaced vector is equal to the magnitude as the original vector. More generally, suppose that the *scalar product* of two arbitrary vectors, A^i and B^j, is invariant under parallel transport. Then the Christoffel symbols are immediately and uniquely defined in terms of the metric tensor, g_{ij}. Thus if we require that

$$g_{ij}(x + dx)DA^i(x + dx)DB^j(x + dx) = g_{ij}(x)A^i(x)B^j(x) \tag{172}$$

for arbitrary A^i and B^j, then it follows (if we assume that the connections are symmetric in their lower indices) that

$$\Gamma^i_{jk} = \frac{1}{2}g^{il}(g_{lj,k} + g_{lk,j} - g_{jk,l}) \tag{173}$$

Exercise: Prove equation (173).

6.8 Geodesic coordinates

In a Riemannian manifold we can always find a coordinate system in which at a given point not only does g_{ij} reduce to diagonal form (i.e. space is locally Minkowskian), but the first derivates of g_{ij} are also zero, i.e. $g_{ij,k} = 0$. In such a coordinate system, equation (173) implies that the Christoffel symbols are all identically zero. We call such a coordinate system a **geodesic coordinate system**.

Exercise: Show that the Christoffel symbols, Γ'^{i}_{jk}, are zero at the point P in the coordinate system defined by

$$x'^{\mu} = x^{\mu} + \frac{1}{2}\Gamma^{\mu}_{\alpha\beta}\,(x^{\alpha} - x^{\alpha}_{P})(x^{\beta} - x^{\beta}_{P})$$

where $\Gamma^{\mu}_{\alpha\beta}$ are the Christoffel symbols evaluated at P in the unprimed coordinate system.