Connections and covariant derivatives

Problem: How do we differentiate a vector or tensor field and preserve the tensor property? There are a number of levels at which this can be discussed. The simplest, as mentioned in the previous section, is to note that the "components" of the obvious thing, namely $\partial V^a/\partial x^b$, do not behave properly under change of basis - they are not the components of a tensor of rank $(1,1)$.

In flat space, there's no problem; we just differentiate. What we're really doing is moving all the vectors parallel to themselves to a single vector space, and doing the differentiation there. In a curved space, this is not possible: it's easily seen (as shown in class) that you can only move a vector (or tensor) parallel to itself along a curve. You cannot propagate it into an open set in such a way that it's parallel unless the space is flat.

With this in mind, we will try to see what we can do, instead of what we can't. Our considerations are purely local, and don't involve the metric tensor initially.

Propagation of a vector along a curve

Suppose $t \to \gamma(t)$ is a curve in some open set U with local coordinates (x^a), so the curve is given parametrically by $t \to x^a(t) : 1 \le a \le n$. Suppose $p, q,$ and r are any three points on the curve. We want to define a map, called **parallel propagation** and denoted Φ_{pq} such that

1. $\Phi_{qp} : T_p(U) \to T_q(U)$ is a vector space isomorphism depending only on the points comprising the curve, and not on the parametrization.
2. $\Phi_{rp} = \Phi_{rq} \circ \Phi_{qp}$ for all points p, q, r on γ.
3. Parallel propagation preserves the scalar product: if $V, W \in T_p(U)$, then $g(V, W) = g(\Phi_{qp}(V), \Phi_{qp}(W))$.

Since the curve is given parametrically, let's use the symbol Φ_t to denote parallel propagation from $T_p = T_{x(0)}$ to $T_{x(t)}$. Let $V_0 \in T_p$, and let $V(t) = \Phi_t V_0$. (We are assuming that parallel propagation exists, and we're going to deduce some equations that must be satisfied.) then

$$\frac{dV}{dt} = \frac{d\Phi_t}{dt} V_0 = \frac{d\Phi_t}{dt} \Phi_t^{-1}(\Phi_t V_0) = \left(\frac{d\Phi_t}{dt}\Phi_t^{-1}\right) V_t, \text{ or}$$

$$\frac{dV}{dt} + \omega V = 0, \text{ where } \omega = -\frac{d\Phi_t}{dt}\Phi_t^{-1}$$

In local coordinates, this will read

$$\frac{dV^a}{dt} + \omega^a_b V^b = 0. \tag{1}$$

Now ω will depend upon the path: it will be a function of $x, \dot{x}, \ddot{x}, \ldots$, all of which are functions of the parameter t. The simplest assumption, and the one that will work for us, is that the matrix ω depends only on $x(t)$ and $\dot{x}(t)$. Moreover, the dependence on $\dot{x}$ is assumed to be *linear*. If the curve is reparametrized with u, we find that ($' = d/du$):

$$\begin{aligned}\frac{dV^a}{du} + \omega^a_b(x, x')V^b &= \frac{dV^a}{dt}\frac{dt}{du} + \omega^a_b(x, \dot{x}\frac{dt}{du})V^b \\ &= \Big(\frac{dV^a}{dt} + \omega^a_b(x, \dot{x})V^b\Big)\frac{dt}{du} \text{ (linearity of } \omega)\end{aligned}$$

That is, under our assumptions, parallel propagation (indicated by the vanishing of these quantities) is independent of the parametrization, and ω is said to define a **linear or affine connection** in U. Specifically, this means that

$$\omega^a_b(x, \dot{x}) = \omega^a_b(x, \dot{x}^c\partial_c) = \omega^a_b(x, \partial_c)\dot{x}^c,$$

and we define the **Christoffel symbols**

$$\Gamma^a_{bc}(x) = \omega^a_b(x, \partial_c). \tag{2}$$

Equation (1) now reads

$$\boxed{\frac{dV^a}{dt} + \Gamma^a_{bc}V^b\dot{x}^c = 0;\ 1 \le a \le n}. \tag{3}$$

This coupled system of n first order ODEs defines the operation of parallel transport along the curve γ. Note that the Christoffel symbols are functions of position only and can be defined in the whole of U in principle. If the curve is given, then $x(t)$ and $\dot{x}(t)$ are known, the Γ's are known, and Eqn (3) is a standard initial value problem with unique solutions - the IC is $V(0) = V_0$. Moreover, the equations are *linear and homogeneous* in V, as they should be. A vector field on γ satisfying eqn (3) is said to be **parallel** along γ. So, given initial conditions V_0 at a point $p \in \gamma$,

- There's a unique solution $V(t)$ to the system (3).
- If $V_0 \in T_pU$, and $q = x(t)$, then $\Phi_{qp}(V_0) = V(t)$, where Φ is the parallel propagator.
- If $V_0, W_0 \in T_pU$, and $V(t), W(t)$ are the corresponding solutions to (3), then $\Phi_{qp}(c_1V_0 + c_2W_0) = c_1\Phi(V_0) + c_2\Phi(W_0)$, so Φ_{qp} is linear, as advertised.

The matrix ω is a matrix of 1-forms, since each "entry" ω^a_b is a linear function on the tangent space. However, ω itself is not a tensor; if we change coordinates to $(\tilde{x}^a)$, then of course we'll require the existence of $\tilde{\omega}$ such that, for the same parallel vector field V[1]

$$\frac{d\tilde{V}}{dt} + \tilde{\omega}\tilde{V} = 0.$$

[1]To take a break from the indices, which are a pain to type, we are using matrix notation here. So V is the vector of components and *not*, as we've previously used this symbol, $V^a\partial_a$.

If $P^a_b = \partial x^a / \partial \tilde{x}^b$ is the change of basis matrix, then in matrix notation we have

$$\begin{aligned}\frac{d\tilde{V}}{dt} &= \frac{d}{dt}(P^{-1}V) \\ &= \dot{P}^{-1}V + P^{-1}\dot{V} \\ &= \dot{P}^{-1}V - P^{-1}\omega V \\ &= -\tilde{\omega}\tilde{V} \\ &= -\tilde{\omega}P^{-1}V\end{aligned}$$

Equating the third and fifth lines above, we get

$$\dot{P}^{-1} - P^{-1}\omega = -\tilde{\omega}P^{-1},$$

or

$$\begin{aligned}\tilde{\omega} &= P^{-1}\omega P - \dot{P}^{-1}P \\ &= P^{-1}\omega P + P^{-1}\dot{P} \qquad (4)\end{aligned}$$

where in the last line, we've used the fact that

$$\frac{d}{dt}(P^{-1}P) = 0 = \dot{P}^{-1}P + P^{-1}\dot{P}.$$

It's the last term of eqn (4) which shows that ω is not a tensor; if you write it out, you'll find that it contains the second derivatives of the coordinates. This is the same correction term we had to add to $\partial_a V^b$ to get a tensor in the previous section; it's the term which vanishes if we only use linear coordinate systems.

□ Definition: For *any* vector field V defined along the curve $x(t)$, the **covariant derivative of V along the curve** is the vector field

$$\frac{DV}{dt} = \frac{dV}{dt} + \omega V.$$

In components, this reads

$$\begin{aligned}\frac{DV^a}{dt} &= \frac{dV^a}{dt} + \omega^a_b V^b \\ &= \frac{dV^a}{dt} + \Gamma^a_{bc}V^b\dot{x}^c.\end{aligned}$$

PROPOSITION: Under a change of coordinates, where P is the change of basis matrix, we have

$$\frac{D\tilde{V}}{dt} = P^{-1}\frac{DV}{dt}.$$

In other words, the covariant derivative of V along γ is another vector field along γ.

PROOF: We have

$$\begin{aligned}
\frac{D\tilde{V}}{dt} &= \frac{d\tilde{V}}{dt} + \tilde{\omega}\tilde{V} \\
&= \frac{d}{dt}(P^{-1}V) + \big(P^{-1}\omega P + P^{-1}dP/dt\big)P^{-1}V \\
&= \frac{dP^{-1}}{dt}V + P^{-1}\frac{dV}{dt} + \big(P^{-1}\omega P - \frac{dP^{-1}}{dt}P\big)P^{-1}V \\
&= P^{-1}\big(\frac{dV}{dt} + \omega V\big) \\
&= P^{-1}\frac{DV}{dt}.
\end{aligned}$$

Of course, if V is a parallel vector field along γ, then $DV/dt = 0$, as above. In any case, we can just compute DV/dt to get another vector field.

♣ Exercise: What is the covariant derivative of ∂_c along the curve? Write it out as a vector, not just its components. Reverting to the usual notation, if $V = V^a\partial_a$, show that

$$\begin{aligned}
\frac{DV}{dt} &= \frac{D}{dt}\big(V^a\partial_a\big) \\
&= \frac{dV^a}{dt}\partial_a + V^a\frac{D\partial_a}{dt}
\end{aligned}$$

Suppose now that V is a vector field defined and smooth in some neighborhood of γ instead of just "on" γ. Then we can expand dV^a/dt and write

$$\begin{aligned}
\frac{DV^a}{dt} &= \frac{dV^a}{dt} + \Gamma^a_{bc}V^b\dot{x}^c \\
&= \frac{\partial V^a}{\partial x^c}\dot{x}^c + \Gamma^a_{bc}V^b\dot{x}^c \\
&= \Big(\frac{\partial V^a}{\partial x^c} + \Gamma^a_{bc}V^b\Big)\dot{x}^c
\end{aligned}$$

Since the left hand side of this is the component of a tensor, and $\dot{x}^c$ is the component of a tensor, we conclude that the quantity in parentheses is likewise a tensor. It has nothing to do with the specific path γ, since it only depends on x.

□ Definition: If V is a vector field, defined in some open set U with local coordinates (x^a), then the quantities

$$\boxed{\nabla_a V^b = \frac{\partial V^b}{\partial x_a} + \Gamma^b_{ca}V^c} \tag{5}$$

are the components of a tensor field of rank $(1,1)$ called the **covariant derivative** of V. Putting in the basis elements, we have

$$\nabla V = \nabla_a V^b dx^a \otimes \partial_b \tag{6}$$

We can extend the covariant derivative to operate on tensors of any rank by requiring that it be appropriately linear, that it satisfy Leibniz's rule, and that it commute with contraction:

☐ Definition: A **covariant derivative** is an operator ∇ on tensor fields which satisfies the following conditions:

1. If T is of rank (r,s), then ∇T is of rank $(r,s+1)$; the covariant rank increases by 1.
2. For any function f, $\nabla(f) = df$. $\left[\nabla_a(f) = \partial_a(f)\right]$
3. For any function f and tensor T, $\nabla(fT) = df \otimes T + f\nabla T$. $\left[\nabla_a(fT^{\cdots}_{\cdots}) = \partial_a(f)T^{\cdots}_{\cdots} + f\nabla_a T^{\cdots}_{\cdots}\right]$.
4. More generally, for any tensors S and T, $\nabla(S \otimes T) = \nabla S \otimes T + S \otimes \nabla T$.

Assuming we know how to differentiate vector fields, we can use the above to determine the covariant derivative of a one form ϕ. Consider the tensor field $\phi \otimes V$ with components $\phi_b V^c$. From (2) and (3) above, we have

$$\begin{aligned}\nabla_a(\phi_b V^c) &= (\nabla_a \phi_b)V^c + \phi_b(\nabla_a V^c)\\ &= (\nabla_a \phi_b)V^c + \phi_b\big(\partial_a V^c + \Gamma^c_{da}V^d\big)\end{aligned}$$

Contracting on b and c, this reads

$$\nabla_a(\phi_b V^b) = (\nabla_a \phi_b)V^b + \phi_b\big(\partial_a V^b + \Gamma^b_{da}V^d\big).$$

But now the left hand side is just a function and by (2) we have $\nabla_a = \partial_a$ on functions. The whole thing now becomes

$$\begin{aligned}\nabla_a(\phi_b V^b) &= (\partial_a \phi_b)V^b + \phi_b(\partial_a V^b)\\ &= (\nabla_a \phi_b)V^b + \phi_b(\partial_a V^b) + \phi_b\Gamma^b_{da}V^d\\ &= (\nabla_a \phi_b)V^b + \phi_b(\partial_a V^b) + \phi_d\Gamma^d_{ba}V^b,\end{aligned}$$

where we have switched two indices in the last term. Equating the first and third lines of this, cancelling the two identical terms and rearranging gives

$$\Big(\nabla_a\phi_b - \big(\partial_a\phi_b - \Gamma^d_{ba}\phi_d\big)\Big)V^b = 0.$$

Since this holds for any vector field V, the quantity in parentheses must vanish, and we obtain

$$\boxed{\nabla_a\phi_b = \partial_a\phi_b - \Gamma^d_{ba}\phi_d} \tag{7}$$

So now we know how to handle both upper and lower indices, and this means we can take the derivatives of tensor fields of any rank. See the exercises below.

Note: The formulas are complicated by the Christoffel symbols because the standard bases for TU and T^*U are not parallel: As an example, for a one-form ϕ, we can also write

$$\begin{aligned}\nabla\phi &= \nabla(\phi_c dx^c)\\ &= d\phi_c \otimes dx^c + \phi_c \nabla(dx^c) \text{ (from (3) in the def'n)}\\ &= \partial_a \phi_c dx^a \otimes dx^c + \phi_c \nabla(dx^c)\end{aligned}$$

To compute $\nabla(dx^c) = \mu_a dx^a$, where $\mu_a = \delta^c_a$, we use (7) and find that

$$\nabla_b \mu_a = \nabla_b \delta^c_a - \Gamma^d_{ab}\delta^c_d = -\Gamma^c_{ab}.$$

So

$$\nabla(dx^c) = \nabla_b \mu_a dx^b \otimes dx^a = -\Gamma^c_{ab} dx^b \otimes dx^a.$$

Plugging this into the expression above, and massaging the indices, we find

$$\nabla\phi = \big(\partial_b \phi_a - \Gamma^c_{ab}\phi_c\big) dx^b \otimes dx^a.$$

And we see that the complications come from the fact that the covariant derivatives of the basis one-forms $\{dx^a\}$ are not zero.

□ Definition: A tensor field T is **parallel** if $\nabla T = 0$.

♣ Exercise:

1. Compute ∇g, where $g = g_{ab} dx^a \otimes dx^b$ using the same techniques as in the above example. What is $\nabla_c g_{ab}$?
2. Without doing any work, guess the form of $\nabla_a T_{bcd}$.
3. $\nabla_a T^b_c = \partial_a T^b_c + \Gamma^b_{da} T^d_c - \Gamma^d_{ca} T^b_d$.

Riemannian geometry

The alert reader has surely noticed that thus far we've made no use of the metric tensor. So there are no constraints on ∇ other than that it behave reasonably as a differential operator. And correspondingly, there is no unique way to choose a connection on U. But if we have a non-degenerate metric g, then it's natural to ask that parallel propagation respect the scalar product.

□ Definition: (**Notation**) Suppose V is a vector field along some curve γ. If V is smooth, it is always possible to extend V smoothly to a neighborhood of γ. In this case, we can write

$$\frac{DV^b}{dt} = T^a \nabla_a V^b, \text{ where } T = T^a \partial_a$$

is the tangent vector to the curve. See eqn (5) and the derivation above it. We can also denote this, in index free notation, by $\nabla_T V$, where

$$\begin{aligned}\nabla_T V &= T^a(\nabla_a V^b)\partial_b \\ &= \nabla_a V^b dx^a(T)\partial_b\end{aligned}$$

It's clear that the result is independent of the particular smooth extension of V to a neighborhood of the curve, so we'll often use this notation without mentioning the extension specifically. The same is true for tensor fields of any rank.

□ Definition: The connection ∇ is said to be a **metric** connection, or **compatible with the metric** if the scalar product between two parallel-propagated vector fields is constant.

Suppose that $U_0, V_0 \in T_pU$, γ is a curve at p and U, V are parallel along the curve, agreeing with U_0, V_0 at $t = 0$. If the connection is compatible with the metric, the function $g(U, V)$ is constant along γ. So

$$\frac{d}{dt}\big(g(U,V)\big) = (\nabla_T g)(U,V) + g(\nabla_T U, V) + g(U, \nabla_T V) = 0.$$

Since U and V are parallel along γ, $\nabla_T U = \nabla_T V = 0$. At the point p,

$$(\nabla_T g)(p)(U_0, V_0) = 0.$$

But U_0 and V_0 are arbitrary in T_pU. Therefore $\nabla_T g = 0$ at p. But the curve γ is arbitrary, so T is arbitrary and therefore $\nabla g(p) = 0$. Since p is arbitrary as well, we must have $\nabla g = 0$ everywhere in U. That is, the metric is covariantly constant (or parallel). Writing this out in detail,

$$\nabla_a g_{bc} = \partial_a g_{bc} - \Gamma^d_{ba} g_{dc} - \Gamma^d_{ca} g_{bd} = 0 \tag{8}$$

We now make one last assumption, that the connection is symmetric:

□ Definition: ∇ is **symmetric** if, in any coordinate system, $\nabla_{\partial_a}\partial_b = \nabla_{\partial_b}\partial_a$. Or equivalently, the connection is symmetric if the Christoffel symbols are symmetric in the two lower indices: $\Gamma^a_{bc} = \Gamma^a_{cb}$.

If ∇ is symmetric and compatible with the metric, the connection turns out to be *unique*: From Eqn (8), $\nabla_a g_{bc} + \nabla_b g_{ac} - \nabla_c g_{ab} = 0$. Using the symmetry of the connection, this yields

$$2\Gamma^d_{ab} g_{dc} = \partial_a g_{bc} + \partial_b g_{ac} - \partial_c g_{ab},$$

Now multiply both sides by $(1/2)g^{ce}$, obtaining

$$\boxed{\Gamma^e_{ab} = \frac{1}{2} g^{ce}\big(\partial_a g_{bc} + \partial_b g_{ac} - \partial_c g_{ab}\big)} \tag{9}$$

□ Definition: Given the metric, the connection *uniquely* determined by Eqn (9) is called the **Riemannian** or **Levi-Civita** connection. Differential geometry, done with this connection, is called **Riemannian geometry**, although some authors reserve this term for the case of positive-definite metrics, and call it **pseudo-Riemannian geometry** if the metric is Lorentzian.

♣ Exercise: On a sphere S^2 of radius r, the components of the metric tensor in the (θ, ϕ) coordinate system are given by

$$(g_{ab}) = \begin{pmatrix} r^2 & 0 \\ 0 & r^2 \sin^2 \theta \end{pmatrix}.$$

Find the Christoffel symbols, using Eqn (9).