What the Hell is a Connection

Ruben Verresen

April 9, 2014

The basic question is: how do we differentiate a section of a vector bundle? Denote $\pi : E \to X$ and an arbitrary section $s : X \to E$. There is a natural object, the differential of our section:

$$\boxed{(ds)_p : T_p X \to T_{s(p)} E}$$

However we don't want our result to be a tangent vector of E, rather we want it to be a vector of E itself (e.g. in the specific case of the tangent bundle $\pi : TX \to X$, where our section is a (tangent) vector field, we want $D_V s$ to be a tangent vector of X again!) Hence we want

$$\boxed{Ds : T_p X \to E_p}$$

(notation: $(Ds)(V)$ is usually written $D_V s$) Written differently, we want $D_V : C^\infty(E) \to C^\infty(E)$, our ("covariant") derivative of sections.

To get such a "Ds" from "ds", the missing link is a map $T_{s(p)}E \to E_p$. Later we will see that mapping **is** our connection! Pictorially, if we take $X = S^1$ and E to be the trivial rank one vector bundle:

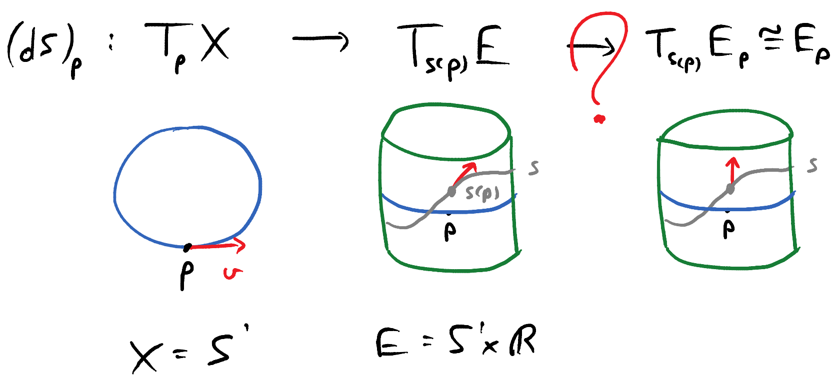

Figure 1: What we *have* (i.e. ds) and what we *want* (i.e. Ds)

Based on this picture, it seems like this should be straightforward: we see we just want the projection onto the fiber! This is misleading however. That's because we can see a natural metric on our bundle, and indeed for a given metric there is a natural connection of sorts (given by this orthogonal projection), called the Levi-Civita connection. But more generally there is no natural choice!

To see what's the tricky part, let's write it out explicitly in coordinates. Take coordinates (x^i) of X. Then if we write vectors of E_p as $v = a^j e_j$ we have coordinates (x^i, a^j) on E. Recall the general formula

$$\text{“} \quad f : X \to Y \qquad \text{then} \qquad df = \frac{\partial f^j}{\partial x^i} \frac{\partial}{\partial y^j} \otimes dx^i \quad \text{”}$$

Applying it to our case gives

$$ds = \frac{\partial x^j}{\partial x^i} \frac{\partial}{\partial x^j} \otimes dx^i + \frac{\partial s^j}{\partial x^i} \frac{\partial}{\partial a^j} \otimes dx^i$$

It *looks* like the first term is just 1 (since $\frac{\partial x^j}{\partial x^i} = \delta^j_i$), in which case our natural choice for D is the second term, which indeed likes wholly in the fiber E_p. This is wrong. Let's be more careful and specify the relevant points at which the objects are defined (in coordinates):

$$(ds)_p = \left(\frac{\partial}{\partial x^i}\right)_{(x,a)} \otimes \left(dx^i\right)_x + \frac{\partial s^j}{\partial x^i} \left(\frac{\partial}{\partial a^j}\right)_{(x,a)} \otimes \left(dx^i\right)_x \tag{1}$$

and realize that we can't meaningfully identify

$$\left(\frac{\partial}{\partial x^j}\right)_{(x,a)} \neq \left(\frac{\partial}{\partial x^j}\right)_x \qquad \left(\text{but } \left(\frac{\partial}{\partial x^j}\right)_{(x,0)} = \left(\frac{\partial}{\partial x^j}\right)_x\right)$$

Let's again explore this pictorially, now taking $X = \mathbb{R}$ and $E = \mathbb{R}^2$: the vertical lines are our fibers. Remember that a^j are our fiber coordinates with respect to e_j, which is a basis that obviously depends on x! Graphically $\partial/\partial x$ corresponds to the direction you move in if you fix the a coordinates. This direction need not be horizontal and depends crucially on $e_j(x)$!

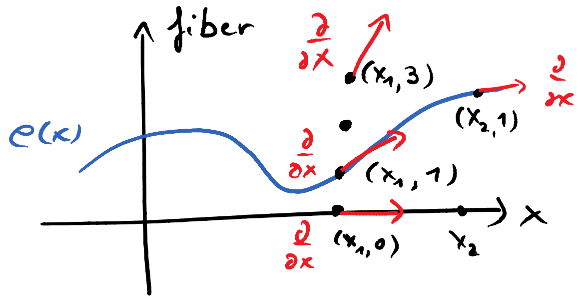

Figure 2: Demonstrating how $\partial/\partial x$ depends on the choice of our basis $e(x)$

Hence we see there's no canonical way of getting rid of the first term in (1). **One approach** is: we choose a certain basis $\{e_j(x)\}$ and *define* it to be flat, i.e. in the picture above we say $e(x)$ is by definition a constant section. This comes down to: ignore the first term in (1) *in this specific basis* (or said differently, *in this specific trivialization*, since choosing a (local) basis comes down to choosing a (local) trivialization of our vector bundle). I.e. define

$$Ds : T_pX \to E_p \text{ given by } \frac{\partial s^j}{\partial x^i} \frac{\partial}{\partial a^j} \otimes dx^i \text{ in the basis/trivialization } \{e_j\}$$

This is one approach, but it's not the most general. In fact, it turns out that all such connections are flat! Indeed, if we interpret flatness as "parallel transport around all loops is trivial", then the above will have that feature. In the above case this is not clear, since all vector bundles on 1-dimensional spaces are flat. Hence let's look at $X = \mathbb{R}^2$ and $E = \mathbb{R}^3$. Now the choice of $e(x)$ is a (non-vanishing) function $\mathbb{R}^2 \to \mathbb{R}$.

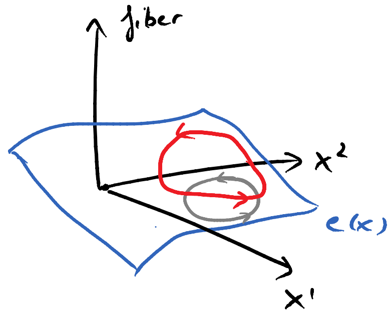

Figure 3: The red loop is a parallelly transported vector along the gray loop

Parallel transport around a loop comes down to just traversing the loop on $e(x)$ as shown (if we start on[1] $e(x)$). The reason for this is simple: $e(x)$ is by definition flat, so it tells us exactly how keep a vector constant. But that means we always end up where we started, this means there is no curvature! (Note this shows that if E is trivial, such that there is a global basis, we can build a flat connection.)

The most **general** thing we can do is simply say: "we project $\frac{\partial}{\partial x^i}$ onto our fiber *one way or another*, denoted by

$$\frac{\partial}{\partial x^i} \mapsto \theta_i^j e_j \quad "$$

This θ_i^j is a function of x^i and a^j. Given the vector structure of our fibers, it is natural[2] to impose

$$\theta_i^j(x, a) = A_{ik}^j(x) a^k$$

This A is called our **connection**. Its role is the projection[3] we needed on page 1:

$$\boxed{A : T_{s(p)}E \to E_p \qquad \text{in coordinates: } A = A_{ik}^j(x)\, a^k e_j \otimes (dx^i)_{(x,a)} + e_j \otimes da^j} \tag{2}$$

In other words, A projects the first term in (1) onto E_p as follows:

$$\left(\frac{\partial}{\partial x^i}\right)_{(x,a)} \otimes \left(dx^i\right)_x \mapsto A_{ik}^j a^k e_j \otimes \left(dx^i\right)_x$$

Note that A contains the second term $e_j \otimes da^j$ to canonically project the second term of (1). In conclusion, from (1) and this projection, we get that our covariant derivative is

$$\boxed{D = A \circ d : s^j e_j \mapsto \left(\partial_i s^j + A_{ik}^j s^k\right) e_j \otimes dx^i}$$

This is sometimes abbreviated as $D = D_A = d_A = d + A$. Be **careful** with this: this is not the d in the ds from before, in fact it's the second term of ds in (1).

[1] Of course we can start *anywhere*, it just becomes a bit harder to draw neatly.
[2] This makes the space of flat sections a vector space.
[3] Its kernal is usually called the "horizontal subspace".

A **second interpretation** of the connection, clearer in the notation $d_A = d + A$, is the deviation of s being constant if we move in a certain direction. More precisely, A_i (supressing the fiber indices) is a linear operator on the fiber telling us "if we move in the direction of increasing x^i, then if I start with a vector s, the deviation from being constant is $A_i s$".

The **first interpretation**, where we see A as being a projection operator, makes the transformation property of A^i_{jk} natural! Indeed, suppose we change trivialization/basis $\{e_i\}$ such that

$$\tilde{a} = \psi a \qquad \tilde{e} = e\psi^{-1} \qquad \big(\text{ s.t. } \tilde{a}^i \tilde{e}_i = \tilde{e} \cdot \tilde{a} = e \cdot a\big)$$

It's handy if we rewrite the definition (2) in this suggestive matrix notation:

$$\begin{aligned} A &= e\,(A_i dx^i)\; a + e \cdot da \\ &= \tilde{e}\psi\,(A_i dx^i)\;\psi^{-1}\tilde{a} + \tilde{e}\psi \cdot d(\psi^{-1}\tilde{a}) \\ &= \tilde{e}\,\big(\,\psi(A_i dx^i)\psi^{-1} + \psi d\psi^{-1}\,\big)\;\tilde{a} + \tilde{e} \cdot d\tilde{a} \end{aligned}$$

This is the well-known **gauge transformation**[4]

$$\boxed{\tilde{A}_i = \psi A_i \psi^{-1} + \psi \partial_i \psi^{-1}} \tag{3}$$

Two comments about this transformation:

1. We implicitly assumed dx^i is the same object whether we use coordinates (x^i, a^j) or $(x^i, \tilde{a}^j)$, but this is **not** as trivial as it seems! The whole reason we had to bother with connections is because $\partial/\partial x^i$ depends on our choice of basis $\{e_i\}$, but as it turns out, dx^i does not :) Mathematically:

$$\left(\frac{\partial}{\partial x^i}\right)_{\text{old}} = \frac{\partial x^j}{\partial x^i}\left(\frac{\partial}{\partial x^j}\right)_{\text{new}} + \frac{\partial \tilde{a}^k}{\partial x^i}\frac{\partial}{\partial \tilde{a}^k} = \left(\frac{\partial}{\partial x^j}\right)_{\text{new}} + \frac{\partial \psi^k_l}{\partial x^i} a^l \frac{\partial}{\partial \tilde{a}^k}$$

$$\boxed{(d\tilde{x}^i)_{\text{new}}} = \frac{\partial \tilde{x}^i}{\partial x^j}(dx^j)_{\text{old}} + \frac{\partial \tilde{x}^i}{\partial a^k} d\tilde{a}^k = \boxed{(dx^j)_{\text{old}}}$$

Graphically:

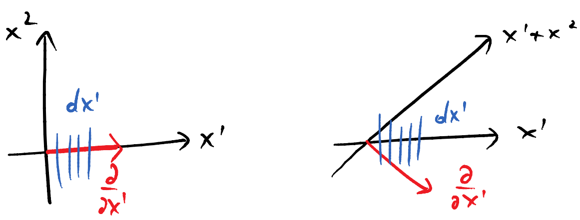

Figure 4: The way $\partial/\partial x$ and dx transform when changing *other* coordinates

[4]Strictly speaking, this is a passive transformation: we change coordinates. In physics we also assume that our physical quantities should be invariant under the active transformation where we *fix* our coordinates but choose a *different* connection $\tilde{A}$ such that its coordinates are numerically the same as those of A in a *different* coordinate system, for which we use (3). Hence whereas gauge invariance is usually seen as a sort of "independence of coordinates" it's saying exactly the opposite: our physical laws *only* use coordinates, hence even two objects that are mathematically different but have coordinate expressions that can be related by coordinate transformations should be seen as the "same" on a physical level.

2. In the case of the tangent bundle $\pi : TX \to X$ we usually do *both* a change of local trivialization (which $\frac{\partial}{\partial x} \mapsto \frac{\partial}{\partial \tilde{x}}$ on the fiber) *and* a coordinate transformation (which $x \mapsto \tilde{x}$ on the base space). This is not *necessary* (one can use old coordinates on the manifold with new coordinates on the tangent space) but it is *natural*, since then we're always using a "coordinate basis" on the tangent space. If we do both, (3) gets an extra transformation factor (which indeed appears for the Christoffel symbol in GR).

Finally, note that there are some other definitions of connections. Firstly, physicists like to take (3) as a definition. And in mathematics one sometimes uses the **axiomatic** definition where we define the covariant derivative as something $D : \Omega^0(E) \to \Omega^1(E)$ that satisfies the Leibniz rule

$$D(fs) = (df)s + fDs$$

From our perspective, we've proven such objects exist, and we started at the natural pre-existing notion of the differential of a section, and then realized the missing object was a projection operator taking the tangent spaces of E to the fibers of E. The axiomatic definition is more amenable for the generalization to $\Omega^p(E)$, where we define

$$D(s\omega) = (Ds) \wedge \omega + sd\omega \qquad s \in \Omega^0(E), \omega \in \Omega^p(\mathbb{R})$$

Again we don't need the axiomatic definition, and we can just define

$$D\omega = d_A\omega = d\omega + A \wedge \omega \qquad \omega \in \Omega^p(E)$$

(one can indeed show this is equivalent). The same **disclaimer** is appropriate: this d is not the usual differential, but rather as before defined by its coordinates

$$d\omega = \partial_i \omega_I \, dx^i \wedge dx^I$$

Note that still $d^2\omega = 0$.

This generalization can be seen as an application of what we have done above: if we have a projection operator A as above, it induces a natural projection operator (from tangent space to fiber) for the vector bundles $\Omega^p(E)$, and that approach is again equivalent to the one described just now. (Am I sure? I might be confusing differentials and exterior derivatives here! Not the same effect, no anti-symmetry in the former!

EDIT: in the fact the "d" in "$d_A = d + A$" can *also* be seen as a proper exterior derivative. This has to do with the fact that on a local trivialization $E \cong X \times \mathbb{R}^k$ we can re-interpret our section as (locally) $s : X \to \mathbb{R}^k$. It is an exterior derivative on this object. For more info, see the blog post.