This article is adapted from “Derivatives of Tensors and the Affine Connection” uploaded by Andrew Dotson, 1 May 2019, https://www.youtube.com/watch?v=NseyJQVpv4U.

We said that there’s a problem if, in taking the derivative of a tensor in non-Cartesian coordinates, we only take the derivative of the tensor’s components without considering basis vectors. The problem is that the result is not a tensor which, in turn, is a problem for—say in physics—where the laws (and the equations that represent them) are known to be invariant in all coordinate systems. Here, we will show exactly how taking the derivative without considering basis vectors creates a problem, and from this, get an idea about how we might go about fixing it.

We’ve seen that the defining feature of tensors is that they are invariant under coordinate transformations. Therefore, we’ll illustrate the problem by showing that taking the derivative in “the normal way” doesn’t yield a tensor.

We’ll start with a position vector, $X^{'i}$. We know that its transformation equation is:

$$X^{'i} = \frac{\partial X^{'i}}{\partial x^{j}} X^{i}$$

Now let’s take the time derivative of $X^{'i}$ (i.e., create a velocity 4-vector):

$$\frac{dX^{'i}}{dt} = \frac{d}{dt}\left(\frac{\partial X^{'i}}{\partial x^{j}} X^{i}\right)$$

Our vectors are not just constants; they are functions of time. Thus, we’ll need to use the chain rule and the product rule. Recall that the chain rule goes like this: $\dfrac{df}{dt} = \dfrac{df}{dx^i}\dfrac{dx^i}{dt}$ and the product rule goes like this: $\dfrac{d(fg)}{dx} = f\dfrac{dg}{dx} + g\dfrac{df}{dx}$. Using these rules, we have:

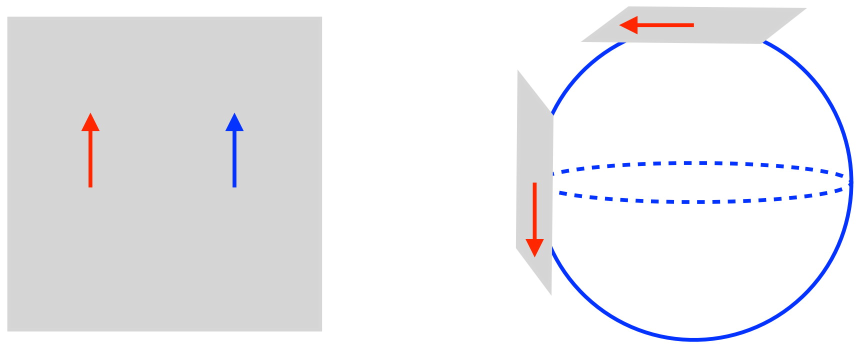
$$\begin{aligned}\frac{dX'^i}{dt} &= \frac{\partial X'^i}{dx^j}\frac{dx^j}{dt} + X^i\frac{d}{dt}\left(\frac{\partial X'^i}{dx^j}\right)\\ &= \frac{\partial X'^i}{dx^j}\frac{dx^j}{dt} + X^i\frac{\partial^2 X'^i}{\partial x^j \partial^k}\frac{dx^k}{dt}\end{aligned}$$

$\dfrac{dX'^i}{dt} = \dfrac{\partial X'^i}{dx^j}\dfrac{dx^j}{dt}$ should be the correct transformation to maintain the tensor status of this derivative, but when we actually do the math, we have this extra term $X^i\dfrac{\partial^2 X'^i}{\partial x^j \partial^k}\dfrac{dx^k}{dt}$. Under some transformations, this term becomes zero and $\dfrac{dX'^i}{dt} = \dfrac{\partial X'^i}{dx^j}\dfrac{dx^j}{dt}$ is, in fact, the correct transformation rule to make the derivative a tensor. For example, if we apply a rotation in Euclidean space, the extra term, indeed, becomes zero and the derivative remains a tensor:

$X' = x\cos\theta + y\cos\theta$ therefore

$\dfrac{\partial X'}{\partial x} = \cos\theta, \quad \dfrac{\partial X'}{\partial y} = \sin\theta$

But $\cos\theta$ and $\sin\theta$ are constants. Their second derivative is zero (because the derivative of a constant is zero). Thus, under such a rotational transformation, the term $X^i\dfrac{\partial^2 X'^i}{\partial x^j \partial^k}\dfrac{dx^k}{dt}$ is zero and the derivative transforms as a tensor. However, under general coordinate transformations, this is often not the case.



a *b*

Figure 1

An example would be curved space (like the surface of a sphere). To evaluate the change in a vector—like a velocity vector that lies in the tangent plane to space in which it dwell—we “move” (i.e., parallel transport) from one point to another and compare. In a space with Cartesian coordinates (figure 1a), this is no problem because the vectors we’re comparing at two points remain in the same vector space at those two points. But in curved coordinates, like on the surface of a sphere (figure 1b), when we parallel transport the vectors and compare them, they wind up in different vector spaces. So how do we compare them? This suggests that we need some method to connect the spaces.

To see how this might be done, consider an observer B jumping out of a plane holding a book and an observer A on the ground. Observer B, in free-fall, thinks he and the book are at rest. Thus, in his frame, he sees no acceleration:

$$\frac{d^2X^\alpha}{d\tau^2} = 0$$

where

X^α is the falling observer’s position vector
τ is proper time

X^α, in turn, is a function of coordinates x^ν, the coordinates of Observer A (i.e., $X^\alpha(x^\nu)$ so, in calculating the derivative above, we need to use the product and chain rule:

$$\frac{d}{dt}\left(\frac{dX^\alpha}{d\tau}\right) = \frac{d}{dt}\left(\frac{dX^\alpha}{dx^\mu}\frac{dx^\mu}{d\tau}\right) = \frac{dX^\alpha}{dx^\mu}\frac{d^2x^\mu}{d\tau^2} + \frac{d^2X^\alpha}{dx^\mu dx^\nu}\frac{dx^\nu}{d\tau}\frac{dx^\mu}{d\tau} = 0$$

We wish to isolate the $\dfrac{d^2x^\mu}{d\tau^2}$ term which would be the acceleration at which Observer A sees Observer B and the book accelerating. To do this, we use the following relationship:

$$\frac{\partial x^\lambda}{\partial X^\alpha}\frac{\partial X^\alpha}{\partial x^\mu} = \delta^\lambda_\mu$$

Specifically, we multiply the $\dfrac{dX^\alpha}{dx^\mu}\dfrac{d^2x^\mu}{d\tau^2}$ term by $\dfrac{\partial x^\lambda}{\partial X^\alpha}$. We get:

$$\begin{aligned}\frac{d}{dt}\left(\frac{dX^\alpha}{d\tau}\right) &= \frac{dX^\alpha}{dx^\mu}\frac{d^2x^\mu}{d\tau^2} + \frac{d^2X^\alpha}{dx^\mu dx^\nu}\frac{dx^\nu}{d\tau}\frac{dx^\mu}{d\tau} = 0\\ &= \frac{\partial x^\lambda}{\partial X^\alpha}\frac{dX^\alpha}{dx^\mu}\frac{d^2x^\mu}{d\tau^2} + \frac{d^2X^\alpha}{dx^\mu dx^\nu}\frac{dx^\nu}{d\tau}\frac{dx^\mu}{d\tau} = 0\\ &= \delta^\lambda_\mu\frac{d^2x^\mu}{d\tau^2} + \underbrace{\frac{\partial x^\lambda}{\partial X^\alpha}\frac{d^2X^\alpha}{dx^\mu dx^\nu}}_{\Gamma^\lambda_{\mu\nu}}\frac{dx^\nu}{d\tau}\frac{dx^\mu}{d\tau} = 0\end{aligned}$$

We'll give the term enclosed by the underbrace, $\Gamma^\lambda_{\mu\nu}$, a name: the affine connection. The delta function changes the term $\dfrac{d^2x^\mu}{d\tau^2}$ to $\dfrac{d^2x^\lambda}{d\tau^2}$ and we are left with:

$$\frac{d^2x^\lambda}{d\tau^2} + \Gamma^\lambda_{\mu\nu}\frac{dx^\nu}{d\tau}\frac{dx^\mu}{d\tau} = 0$$

So we see again the second derivative equations that represent motion in the free-fall coordinate system and the ground coordinate system differ (i.e., are not invariant under coordinate transformations, and therefore, aren't tensors). We also note that the term $\Gamma^\lambda_{\mu\nu}\frac{dx^\nu}{d\tau}\frac{dx^\mu}{d\tau}$ relates (connects) the two second derivatives and the object $\Gamma^\lambda_{\mu\nu}$ plays an important role in this connection.

As an aside, in our example, notice the similarity between our equation

$$\frac{d^2x^\lambda}{d\tau^2} + \Gamma^\lambda_{\mu\nu}\frac{dx^\nu}{d\tau}\frac{dx^\mu}{d\tau} = 0$$

and the equation that describes a free-falling object in classical mechanics

$$\frac{d^2x^\lambda}{d\tau^2} - g = 0$$

suggesting that this affine connection, in this case, has something to do with gravity.

At any rate, the main goal of this discussion was to point out that 1) there's a problem with trying to take a derivative of components alone in differing coordinate systems (namely, the resulting mathematical objects are not tensors i.e., are not invariant) and 2) the affine connection may be part of the solution.