Geometric definitition of covariant derivative

In introducing the concept of the covariant derivative in courses in General relativity, the procedure is often to introduct the connection $\Gamma^i_j k$ as an extra object into the list of objects defined in the spacetime of interest. As in Wald [1] or Carroll[2], this object is assumed to have a variety of properties. If one demands that these symbols are symmetric in jk and that the metric has a zero covariant derivative, one gets the usual Levi-Civita symbols. However, the set of conditions seem somewhat artifical, and students are often confused as to why they those assumptions are made. In particular it seems as though theories with torsion (anti-symmetric parts to the connection) are just as reasonable and natural as the usual definition.

I want to argue that the usual Christoffel symbols are in fact priviledged, and have a natural geometric definition. Ie, any theory with a metric already has, for geometric reasons, the symmetric connection. This geometric definition of the derivative can be proven to obey the various conditions rather than postulating them.

The derivative of functions along curves is natural and in fact forms the basis for the definiton of tangent and cotangent vectors. If we have functions f(p) from the points of a space to real numbers, and curves $\gamma(\lambda)$ of functions from the real line to points of the space, one can define tangent vectors at a point of the space p as equivalance classes of curves through the point such that two curves $\gamma_1(\lambda)$ and $\gamma_2(\lambda)$ have the same tangent vector iff the derivative $\frac{d}{d\lambda}f(\gamma_1(\lambda)) = \frac{d}{d\lambda}f(\gamma_2(\lambda))$ for all functions f. (Similarly, the cotangent vector, or gradient, is the equivalence class of functions such that $\frac{d}{d\lambda}f_1(\gamma(\lambda)) = \frac{d}{d\lambda}f_2(\gamma(\lambda))$ for all curves $\gamma(\lambda)$) I will use Penrose's abstract index notation, so that V^A simply denotes a tangent vector (by the upper capital roman index) with the name "V".

Given that one has a set of tangent vectors, V^A defined at the various points along a curve $\gamma(\lambda)$, one would like to define the derivative of this tangent vector along the curve. But tangent vectors at two separate points of the spacetime cannot be added, and thus the usual definition of a derivative

$$D_\gamma V^A = \lim_{\epsilon \to 0} \frac{V(\gamma(\lambda+\epsilon)) - V^A(\gamma(\lambda))}{\epsilon} \tag{1}$$

makes no sense since the difference of tangent vectors at two separate points makes no sense.

However, if the space has a metric, g^{AB}, one can define geodesics between two points as an extremal length curve between the two points. Using the definition of the length of the curve

$$s = \int_{\lambda_1}^{\lambda_2} \sqrt{|g_{AB}U^AU^B|}d\lambda \tag{2}$$

where $\gamma(\lambda_1) = p_1$ and $\gamma(\lambda_2) = p_2$, and demanding that the curve be extremal, one gets, in a coordinate system, that the curve must obey the equation

$$\frac{d^2x^i}{ds^2} + g^{il}\left[\partial_j g_{kl} - \frac{1}{2}\partial_l g_{jk}\right]\frac{dx^j}{ds}\frac{dx^k}{ds} \tag{3}$$

Only the symmetric part in j, k of those derivatives of the metric contribute, so we define the usual Christoffel symbols by

$$\Gamma^i_{jk} = \frac{1}{2}g^{il}\left[\partial_j g_{kl} + \partial_k g_{jl} - \partial_l g_{jk}\right] \tag{4}$$

That this expression for Γ occurs in the equations for the geodesic follow purely from the geometric requirement that the geodesic be an extremal curve between the two point and not as some generic definition of parallelism for vectors. The definition of the geodesic and of these symbols does not depend on any other structure of the spacetime other than the existence of a metric.

Define the derivative of a vector field geometrically. Consider the curve $\gamma(\lambda)$ such that for each point λ we have a vector $V^A(\lambda)$ defined at the point $\gamma(\lambda)$. Without loss of generality, I will be interested in defining the derivative of this set of vectors along the curve γ at the point $p = \gamma(0)$. choose for each λ near 0 an arbitrary one of the infinite number of curves in the equivalence class which have $V^A(\lambda)$ as its tangent vector. Ie, for each λ choose a curve $\gamma_\lambda(\mu)$ such $\gamma_\lambda(0) = \gamma(\lambda)$ and that $V^A(\lambda)$ is its tangent vector at that point. For each lambda construct another curve $\tilde{\gamma}_\lambda(\mu)$ through $\gamma(0)$ by the following procedure:

For each value of λ and μ join the point $p = \gamma(0)$ to the point $\gamma_\lambda(\mu)$ by a geodesic. (if there is more than one, choose the one that goes to zero length as ϵ and μ go to zero). Now take the half way point along this geodesic according the length of the this curve. Join the point $\gamma(\lambda)$ to this midpoint by a geodesic, and extend this geodesic to double its length. Define that end point as the point along a curve $\tilde{\gamma}_\lambda(\mu)$. As μ goes to zero, this curve goes to the point $p = \gamma(0)$, since the geodesic from p to $\gamma_\lambda(0) = \gamma(\lambda)$ is the same as the geodesic from $\gamma_\lambda(0)$to the midpoint of tha first geodesic and doubling its length brings us back to the original point p.

The tangent vectors to all of the curves $\tilde{\gamma}_\lambda$ at $\mu = 0$ are thus all defined at the same point p and one can define the derivative of these sets of tangent vectors all lying at the same

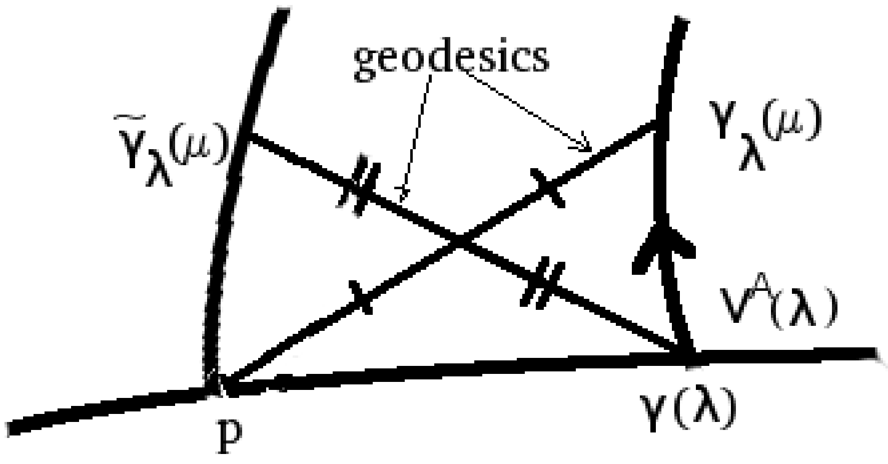

FIG. 1: The translation of the curve γ_λ which runs through the point $\gamma(\lambda)$ to the curve $\tilde{\gamma}_\lambda$ which runs through the point p by constructing geodesic diagonals which intersect at their midpoints. One cunstructs these diagonals for each point $\gamma_\lambda(\mu)$ to generate the point $\tilde{\gamma}_\lambda(\mu)$. The tangent vector to $\tilde{\gamma}_\lambda$ at P is then the geometric transport of the tangent vector to γ_λ at the point $\gamma(\lambda)$.

point p. This derivative will be the covariant derivative of the tangent vectors V^A defined along the curve $\gamma(\lambda)$.

Ie, if we define

$$P_\lambda V^A = \left(\frac{\partial}{\tilde{\gamma}_\lambda}\right)^A \tag{5}$$

the tangent vector to curve $\tilde{\gamma}_\lambda$ at the point p then, the geometric derivative of V^A along the curve γ is

$$D_\gamma V^A = \lim_{\lambda\to 0} \frac{P_\lambda V^A - V^A(p)}{\lambda} \tag{6}$$

which is well defined since $P_\lambda V^A$ and $V^A(p)$ are vectors defined at the same point p and can thus be subtracted, and divided by λ.

To evaluate the components of this derivative in some coordinate system, we impliment the above procedure in that coordinate system. Without loss of generality, we can define the coordinates all to have value 0 at the point p. Now define

$$x^i(\lambda,\mu) = x^i(\gamma_\lambda(\mu)) \tag{7}$$

the value of the coordinates along the curve γ_λ at parameter μ. The geodesic equations are

$$\frac{d^2x^i}{ds^2} = -\Gamma^i_{jk}(x(s))\frac{dx^j}{ds}\frac{dx^k}{ds} \tag{8}$$

where s is the pathlength parameter. Solving for the geodesic from p to $\gamma_\lambda(\mu)$ and retaing only terms to quadratic order in s we get

$$x^i(\lambda,\mu) \approx W^i s - \frac{1}{2}\Gamma^i_{jk}(0)W^jW^ks^2 \tag{9}$$

for some components W^i and the midpoint lies at half that value for s, Keeping terms only to the lowest order we get

$$W^i s \approx x^i(\lambda,\mu) + \frac{1}{2}\Gamma^i_{jk}(0)x^j(\lambda,\mu)x^k(\lambda,\mu) + O(s^3) \tag{10}$$

and the midpoint will lie at

$$x^i_M \approx \frac{1}{2}(W^i s) - \frac{1}{8}\Gamma^i_{jk}(0)W^jW^ks^2 \tag{11}$$

$$\approx \frac{1}{2}x^i(\lambda,\mu) + \frac{1}{8}\Gamma^i_{jk}(0)x^j(\lambda,\mu)x^k(\lambda,\mu) \tag{12}$$

Now solving for the geodesic from $\gamma(\lambda)$ to the midpoint we get

$$x^i_M = x^i(\lambda,0) + U^i s - \frac{1}{2}\Gamma^i_{jk}U^jU^ks^2 \tag{13}$$

or

$$U^i s = x^i_M - x^i(\lambda,0) + \frac{1}{2}\Gamma^i_{jk}(x^j_M - x^j(\lambda,0))(x^k_M - x^k(\lambda,0)) \tag{14}$$

The point along the curve $\tilde{\gamma}_\lambda(\mu)$ lies twice as far along this curve than does the midpoint, which is

$$\begin{aligned}
\tilde{x}^i(\lambda,\mu) &= x^i(\lambda,0) + 2U^i s - 2\Gamma^i_{jk}U^jU^ks^2 && (15)\\
&= x^i(\lambda,0) && (16)\\
&+ 2\left(x^i_M - x^i(\lambda,0) + \frac{1}{2}\Gamma^i_{jk}(x^j_M - x^j(\lambda,0))(x^k_M - x^k(\lambda,0))\right) && (17)\\
&\quad - 2\Gamma^i_{jk}(x^j_M - x^j(\lambda,0))(x^k_M - x^k(\lambda,0)) && (18)\\
&= 2x^i_M - x^i(\lambda,0) - \Gamma^i_{jk}(x^j_M - x^j(\lambda,0))(x^k_M - x^k(\lambda,0))) && (19)
\end{aligned}$$

This is a set of curves that go through the point p. The components of the tangent vector to these curves is the derivative of the coordinate with respect to μ at $\mu = 0$. The derivative of these vectors with respect to λ is just the derivative of the components, since the basis coordinate tangent vectors do not depend on λ. Thus

$$(D_\gamma V)^i = \partial_\lambda \partial_\mu \tilde{x}^i(\lambda,\mu) \tag{20}$$

$$= 2\partial_\lambda \partial_\mu x^i_M - 2\Gamma^i_{jk}\partial_\mu x^j_M \partial_\lambda (x^k_M - x^k(\lambda,0) \tag{21}$$

But

$$x^i_M = \frac{1}{2}x^i(\lambda,\mu) + \frac{1}{8}\Gamma^i_{jk}x^j(\lambda,\mu)x^k(\lambda,\mu) \tag{22}$$

and we have

$$(D_\gamma V)^i = \partial_\lambda \partial_\mu x^i(\lambda,\mu) + \Gamma^i_j k \partial_\mu x^j \partial_\lambda x^k \tag{23}$$

$$= \partial_\lambda V^i + \Gamma^i_{jk} V^j \partial_\lambda x^k \tag{24}$$

which is the usual metric definition of the covariant derivative.

This derivative is defined in a completely geometric fashion and impliments the intuition that the opposite sides of a parallelogram (defined by its intersecting straight diagonals) are parallel to each other. From this definition, it immediately follows that the torsion is zero. (This arises because we used intersecting diagonals). The choice of the metric covariant derivative is not one arbitrary choice amongst and infinite number of possibilities, but is the connection naturally forced on one purly from the existence of the metric.

One can now extend the definition of the covariant derivative from tangent vectors to other objects. Given any tensor, ie, linear function of tangent vectors, we can define the transported tensor from the point $\gamma(\lambda)$ to $\gamma(0)$ by demanding that the the transported tensor evaluated on the transported tangent vectors is the same as the original tensor on the original tangent vectors. Ie, if we have a tensor $T(V^A, U^A)$ of tangent vectors at the point $\gamma(\lambda)$, we can define the tensor $P_\lambda T$ at the point $\gamma(0)$ by demanding that

$$P_\lambda T(P_\lambda V^A, P_\lambda U^B) = T_\gamma(\lambda)(V^A, U^B) \tag{25}$$

for all V^A, U^B. Since a tensor is completely defined by its evaluation on arbitrary vectors, this will completely define the tensor PT. Since cotangent vectors are tensors with a single tangent vector as argument, this defines the transport of cotangent vectors as well, and the transport of arbitrary tensors and thus of arbitrary tensors.

The derivative of a tensor is then defined in the usual way

$$D_\gamma T_{AB} = \lim_{\lambda\to 0} \frac{P_(\lambda) T_{AB} - T(0)_{AB}}{\lambda} \tag{26}$$

From this definiton one can derive other properties of the covariant derivative of tensors– linearity, product rule, etc– just as one can for functions It is not necessary to postulate them, as is sometimes claimed in the literature.

While one can certainly define other notions of parallelism by adding extra tensors to the definition of the derivative

$$\hat{D}_\gamma V^A = D_\gamma V^A + S^A_B V^B \tag{27}$$

where S may depend on the curve γ (eg, be a tensor function of the tangent vector to the curve), this definion makes clear that that that tensor S defines an additional type of derivative, and that one always has the usual metric derivative defined in the theory as long as one has a metric.

---

[1] R.M. Wald "General Relativity", University Of Chicago Press(1984)

[2] S. Carroll " Spacetime and Geometry: An Introduction to General Relativity", Benjamin Cummings (2003)