General Relativity: Covariant derivation

Gonzalo E. Reyes
Université de Montréal

15 April 2006

Covariant derivation is an operation that comes in several guises. Rather than to try to give the most general definition, we look at some particular examples which will be used in the sequel. The first is not really an example of this operation, but it appears in the formulation of their properties.

Since we want to compare these notions with the corresponding classical ones, we devote the first section to some preliminary classical results.

0.1 Vector fields on R^n

From now on, we assume that M is an n-dimensional manifold, i.e., locally homeomorphic to R^n. Since the notions to be described are local, we assume without loss of generality that $M = R^n$.

Canonical vector fields on R^n

We define the i^{th} *canonical vector field* $\partial/\partial x^i : M \longrightarrow M^D$ by the rule

$$(\partial/\partial x^i)(x^1, \ldots, x^n) = [d \mapsto (x^1, \ldots, (x^i + d), \ldots, x^n)]$$

or, what amounts to the same,

$$(\partial/\partial x^i)|_x = [d \mapsto x + de_i]$$

where $e_i = (0, \ldots, 1, \ldots, 0)$ (the 1 in the i^{th} position) and $x = (x^1, \ldots, x^n)$.

By identifying M^D with $M \times R^n$ we may write

$$(\partial/\partial x^i)|_x = (x, e_i)$$

The point about these canonical vector fields is that for every $x \in M$, $\{\partial/\partial x^1|_x, \ldots, \partial/\partial x^1|_x\}$ constitutes a basis of M_x. Thus

Proposition 1 *Every vector field $X : R^n \longrightarrow (R^n)^D$ can be written uniquely as*

$$X(x) = \sum_{i=1}^{n} v^i(x)\partial/\partial x^i|_x$$

where $v^i : R^n \longrightarrow R$.

In a similar way,

Proposition 2 *Every vector field $V : R \longrightarrow M^D$ along a curve γ can be written uniquely as*

$$V(t) = \sum_{i=1}^{n} v^i(t)\partial/\partial x^i|_{\gamma(t)}$$

0.2 Directional derivative of a function

Let $f : M \longrightarrow R$ be a function and $X \in \mathcal{X}(M)$, i.e., a vector field on M. We define the *directional derivative of* f to be the function $X(f) : M \longrightarrow R$ defined as follows. For $x \in M$, $f \circ X_x : D \longrightarrow R$ can be written as

$$(f \circ X_x)(d) = f(x) + d.X(f)(x)$$

for a unique $X(f)(x) \in R$, by the Kock-Lawvere axiom. It is easy to show that $X(-) : R^M \longrightarrow R^M$ is linear and satisfies Leibniz' rule, i.e.,

$$X(a.f) = a.X(f)$$
$$X(f+g) = X(f) + X(g)$$
$$X(fg) = f.X(g) + g.X(f)$$

Assume that $M = R^n$, $f : R^n \longrightarrow R$ is a function and X is given as

$$X(x) = \sum_{i=1}^{n} v^i(x)\partial/\partial x^i|_x$$

We claim that the directional derivative of f along X is

$$X(f)(x) = \sum_{i=1}^{n} v^i(x)(\partial f/\partial x^i)(x)$$

Indeed, the vector field X can also be written as

$$X_x(d) = x + d(v^1(x), \ldots, v^n(x)) = (x^1 + d.v^1(x), \ldots x^n + d.v^n(x))$$

By Taylor expansion $f(X_x(d)) = f(x) + d\sum_{i=1}^n v^i(x)(\partial f/\partial x^i)(x)$. This implies the conclusion, by the very definition of $X(f)$.

0.3 Covariant derivative of of an E-vector field

Let $p : E \longrightarrow M$ be a vector bundle, ∇ an affine connection on M, $X \in \mathcal{X}(M)$ and $Y : M \longrightarrow E$ an E-vector field.

We define a new E-vector field $\nabla_X Y : M \longrightarrow E$ by the equation

$$h(\nabla_X Y)_m = \nabla(d \mapsto \tilde{X}_m(h+d), Y_{X_{m(h)}})(-h) - Y_m$$

for every $h \in D$, where $\tilde{X} : M \longrightarrow M^{D_2}$ is the canonical extension of M. (The fact that this definition makes sense follows from the Kock-Lawvere axiom applied to the fiber $p^{-1}(m)$.)

Theorem 3 *Let E, ∇ as before, $X_1, X_2 \in \mathcal{X}(M)$ and $Y, Z : M \longrightarrow E$ be two E-vector fields. Furthermore, let $f : M \longrightarrow R$.Then*

1. $\nabla_{X_1 \oplus X_2} Y = \nabla_{X_1} Y + \nabla_{X_2} Y$
2. $\nabla_{f.X} Y = f.\nabla_X Y$
3. $\nabla_X (Y + Z) = \nabla_X Y + \nabla_X Z$
4. $\nabla_X (f.Y) = f.\nabla_X Y + X(f).Y$

where $X(f)$ is the directional derivative of f in the direction X. (+ refers to addition in E, $\oplus$ to addition in M^D.)

Proof: Cf. Mo/Re page 201. Since

$$\begin{cases} (Y.(X_1 \oplus X_2))_m = (Y.X_1) \oplus (Y.X_2)_m \ \text{(the latter } \oplus \text{ in } E^D) \\ (Y.(fX))_m = f(m) \odot (Y.X)_m \ (\odot \text{ in } E^D) \\ ((Y+Z) \odot X)_m = (Y.X)_m + (Z.X)_m \ \text{(the latter + in } E^D) \end{cases}$$

are clear from the linearity properties of the connection map C. To prove 4., we use the geometric interpretation of $\nabla_X Y$. Thus, for every $h \in D$,

$$h.\nabla_X(f.Y)_m = \nabla(t, f(X_m(h))Y_{X(m)})(-h) - g(m)Y_m$$

where $t(d) = \tilde{X}(d+h, m)$.

On the other hand, $f(X_m)(h)) = f(m) + h.X(f)$, by definition of the directional derivative $X(f)$. Using homogeneity,

$$\begin{aligned} h.\nabla_X(f.Y)_m &= (f(m) + hX(f)(m))\nabla(t, Y_{X_m(h)})(-h) - f(m)Y_m \\ &= f(m).h.(\nabla_X Y)_m + hX(f)(m)\nabla(t, Y_{X_m(h)})(-h) \end{aligned}$$

Since $\nabla(t, Y_{X_m(h)})(-h) = Y_m + h(\nabla_X Y)_m$ and $h^2 = 0$, the last term equals

$$h.(f(m)(\nabla_X Y)_m + X(f)(m)Y_m).$$

Since this holds for every $h \in D$, from Kock-Lawvere for E_m we conclude that

$$\nabla_X(fY)_m = f(m)(\nabla_X Y)_m + X(f)(m)Y_m$$

In most of the applications, $E = M^D$

Assume that $M = R^n$ and $E = (R^n)^n$.

Proposition 4

$$(\nabla_{\partial/\partial x^i}\partial/\partial x^j)_x = \sum_{k=1}^{n} \Gamma^k_{ij}(x)\partial/\partial x^k|_x$$

Proof: We recall that $\Gamma^k_{ij}(x) = (\nabla_4((x, e_i).(x, e_j))_k$ i.e., the k^{th} component of $\nabla_4((x, e_i).(x, e_j))$.

By definition of $(\nabla_X Y)_x$,

$$\begin{aligned} h.(\nabla_{\partial/\partial x^i}\partial/\partial x^j)_x &= \nabla([d \mapsto (\widetilde{\partial/\partial x^i})_x(h+d)], (\partial/\partial x^i)\partial/\partial x^j)_{\partial/\partial x^i})(-h) \\ &\quad -(\partial/\partial x^j)_x \end{aligned}$$

We first compute the first term on the right, i.e., $\nabla(t_1, t_2)$ with

$$\begin{cases} t_1(d) = (\widetilde{\partial/\partial x^i})_x(h+d) = (x^1, \ldots, x^i + h, \ldots, x^n) + de_i \\ t_2(d) = (\partial/\partial x^i)\partial/\partial x^j)_{\partial/\partial x^i} = (x^1, \ldots, x^i + h, \ldots, x^n) + de_j \end{cases}$$

Notice that $t_1(0) = t_2(0) = (x^1, \ldots, x^i + h, \ldots, x^n) = x + he_i$

By definition of ∇_4 (cf. "affine connections.pdf" page 3, where ∇_4 is called $\tilde{\nabla}_a$.)

$$\nabla(t_1, t_2)(d_1, d_2) \quad = x + he_i + d_1 e_i + d_2 e_j - d_1 d_2 \nabla_4((t_1(0)), e_i), (t_1(0), e_j))$$

Therefore,

$$\nabla(t_1, t_2)(-h, d) = x + de_j + dh\nabla_4((x + he_i, e_i), (x + he_i, e_j))$$

Letting $g(h) = \nabla_4((x+he_i, e_i), (x+he_i, e_j))$ and developing in Taylor series, $g(h) = g(0) + hg'(0)$. Thus,

$$\nabla(t_1, t_2)(-h, d) = x + de_j + dh\nabla_4((x, e_i), (x, e_j))$$

Returning to the notation $(a, b) = [d \mapsto a + bd]$

$$\nabla(t_1, t_2)(-h) = (x, e_j) + h.(x, \nabla_4((x, e_i), (x, e_j))$$

On the other hand, $(\partial/\partial x^j)_x(d) = x + de_j$ and

$$\begin{array}{rcl} \nabla_4((x, e_i), (x, e_j)) & = & \sum_{k=1}^n e_k \nabla_4((x, e_i), (x, e_j))_k \\ & = & \sum_{k=1}^n \Gamma_{ij}^k(x) \partial/\partial x^k|_x \end{array}$$

Therefore after substraction and cancelling the h's in both members,

$$(\nabla_{\partial/\partial x^i} \partial/\partial x^j)_x = \sum_{k=1}^n \Gamma_{ij}^k(x) \partial/\partial x^k|_x$$

0.4 Covariant derivative of a vector field relative to a curve

Let $Y : R \longrightarrow M^D$ be a vector field and $\gamma : R \longrightarrow M$ a curve. Define a vector field along γ, $\nabla_{\gamma^\bullet} Y$, by the universal validity (in h) of the equation

$$h.(\nabla_{\gamma^\bullet} Y)(t) = \nabla(\gamma^\bullet(t), Y(\gamma(t+h))(-h) - Y(\gamma(t))$$

Proposition 5 *The operation $\nabla_{\gamma^\bullet}$ has the following properties*

(a) $\nabla_{\gamma^\bullet}(Y + Z) = \nabla_{\gamma^\bullet}(Y) + \nabla_{\gamma^\bullet}(Z)$

(b) $\nabla_{\gamma^\bullet} fY = f\circ\gamma \nabla_{\gamma^\bullet} Y + d/dt(f\circ\gamma)Y\circ\gamma$

(c) $\nabla_{\gamma^\bullet} Y = D(Y\circ\gamma)/dt$

Proof: (a) is immediate. As for (b) : a simple computation gives

$$\begin{array}{rcl} h(\nabla_{\gamma^\bullet} fY)(t) & = & \nabla(\gamma^\bullet(t), f(\gamma(t+h))Y(\gamma(t+h))(-h) - f(\gamma(t))Y(\gamma(t) \\ & = & \nabla(\gamma^\bullet(t), (f(\gamma(t)) + hd/dtf(\gamma(t))Y(\gamma(t+h))(-h) \\ & - & f(\gamma(t)Y(\gamma(t) \\ & = & f(\gamma(t)[\nabla(\gamma^\bullet(t), Y(\gamma(t+h)(-h) - Y(\gamma(t))] \\ & + & h\nabla(\gamma^\bullet(t), d/dtf(\gamma(t))Y(\gamma(t+h))(-h) \\ & = & f(\gamma(t))h(\nabla_{\gamma^\bullet} Y)(t) + hd/dtf(\gamma(t))Y(\gamma(t)) \end{array}$$

Canceling the universally quantified h we obtain the conclusion. (To obtain the last equation, we put h in one of the arguments of ∇, obtaining $h^2 = 0$ for this argument and using the definition of ∇).

Finally, (c) is immediate by comparing definitions. In fact, these covariant derivatives are characterized by the equations

$$\begin{cases} h.(DV/dt)(t) = \nabla(\gamma^\bullet(t), V(t+h))(-h) - V(t) \\ h.(\nabla_{\gamma^\bullet} Y)(t) = \nabla(\gamma^\bullet(t), Y(\gamma(t+h))(-h) - Y(\gamma(t)) \end{cases}$$

for all $h \in D$ and they are identical when $V = Y\circ\gamma$.

0.5 Covariant derivative of a curve of vectors

Let $p : E \longrightarrow M$ be a vector bundle and let ∇ be an affine connection on E with connection map C. Let $V : R \longrightarrow E$ be a curve of vectors. The *covariant derivative of* X is the new curve of vectors DV/dt defined by

$$DV/dt(t) = C(d \mapsto V(t+d))$$

Proceeding just as in the case of $\nabla_X Y$, we can show

Proposition 6 *DV/dt is uniquely determined by the following identity*

$$h.DV/dt(t) = \nabla(d \mapsto p\circ\tilde{t}(h+d), V(t+h))(-h) - V(t)$$

for every $h \in D$, where $\tilde{t} : D_2 \longrightarrow E$ is such that $\tilde{t}(\delta) = V(t+\delta)$.

This proposition says that $DV/dt(t)$ is computed by transporting $V(t+h)$ back along $p \circ V$ to the point t, and substracting $V(t)$ from the result.

In most of the applications, we let $E = M^D$.

Proposition 7 *The operation DV/dt has the following properties:*

(a) $D(V+W)/dt = DV/dt + DW/dt$

(b) $D(f.V)/dt = (df/dt)V + f.DV/dt$ *whenever* $f : R \longrightarrow R$

(c) $DV/dt = \nabla_{\gamma^\bullet} Y$ *whenever* $V = Y \circ \gamma$ *locally.*

Proof: Recall that DV/dt is uniquely determined by the following identity

$$h.DV/dt(t) = \nabla(d \mapsto p \circ \tilde{t}(h+d), V(t+h))(-h) - V(t)$$

for every $h \in D$, where $\tilde{t} : D_2 \longrightarrow E$ is such that $\tilde{t}(\delta) = V(t+\delta)$.

[Ad (a)] Immediate from the above characterization and the bilinearity of ∇

[Ad (b)] Use again the above characterization to obtain

$$h(D(f.V)/dt)(t) = \nabla(d \mapsto p \circ \tilde{t}(h+d), f(t+h)V(t+h))(-h) - f(t)V(t)$$

From $f(t+h) = f(t) + hf'(t)$ and bilinearity of ∇,

$$\begin{aligned} h(D(f.V)/dt)(t) &= f(t)(\nabla(d \mapsto p \circ \tilde{t}(h+d), V(t+h))(-h) - V(t)) \\ &+ hf'(t)\nabla(d \mapsto p \circ \tilde{t}(h+d), V(t+h))(-h) \\ &= f(t)hDV/dt(t) + hf'(t)[hDV/dt(t) + V(t)] \\ &= hf(t)DV/dt(t) + hf'(t)V(t) \end{aligned}$$

Canceling the universally quantified h we obtain the desired result.

[Ad (c)] Already proved (cf. proposition 5.)

Proposition 8 $V \mapsto DV/dt$ *is the only operation from vector fields along γ into vector fields along γ with the properties* $(a) - (c)$.

Proof: Write

$$V(t) = \sum_{j=1}^{n} v^j(t)\partial/\partial x^j|_{\gamma(t)}$$

and compute DV/dt using $(a)-(c)$

$$\begin{array}{rcl}(DV/dt)(t) & = & \sum_{j=1}^{n}(D/dt)(v^j(t)\partial/\partial x^j|_{\gamma(t)}) \text{ (a)} \\ & = & \sum_{j=1}^{n}((dv^j/dt)(t)\partial/\partial x^j|_{\gamma(t)} + v^j(t)(D/dt)\partial/\partial x^j|_{\gamma(t)})) \text{ (b)} \\ & = & \sum_{j=1}^{n}((dv^j/dt)(t)\partial/\partial x^j|_{\gamma(t)} + v^j(t)\nabla_{\gamma^\bullet}\partial/\partial x^j) \text{ (c)} \\ & = & \sum_{j=1}^{n}((dv^j/dt)(t)\partial/\partial x^j|_{\gamma(t)} \\ & + & v^j(t)\sum_{i=1}^{n} d\gamma^i/dt\nabla_{\partial/\partial x^i|_{\gamma(t)}}\partial/\partial x^j) \\ & = & \sum_{k=1}^{n}(dv^k/dt + \sum_{i,j=1}^{n}\Gamma_{ij}^k(\gamma(t))(d\gamma^i/dt)v^j(t))\partial/\partial x^k|_{\gamma(t)})\end{array}$$

Conversely, the operation defined by this formula has properties $(a)-(c)$. Notice that for (c) it is enough to require that $V = Y \circ \gamma$ only locally.

Since (a) and (b) follow from the addition rule and product rule for ordinary derivatives, respectively, we only have to prove (c).

Applying our formula for the covariant derivation to $Y = \partial/\partial x^l$ we obtain

$$D/dt\partial/\partial x^l|_{\gamma(t)} = \sum_{i,k}\Gamma_{il}^k(\gamma(t))d\gamma^i/dt\partial/\partial x^l|_{\gamma(t)}$$

But this is precisely $\nabla_{\gamma^\bullet}\partial/\partial x^l$ (cf. proposition 4).

The general case follows from bilinearity of D/dt and $\nabla_{\gamma^\bullet}$ together with Leibniz' law for D/dt and $\nabla_{\gamma^\bullet}$. (cf. proposition 5)

We say that a vector field V along γ *is parallel* along γ iff $DV/dt = 0$ along γ.

Proposition 9 *Let V be a vector field along γ. Then V is parallel along γ iff $V(t+h) = \tau_h(\gamma, V(t))$, for every $t \in R$ and $h \in D$.*

Proof: Recall that DV/dt is characterized by the equation

$$hDV/dt(t) = \nabla(\gamma^\bullet(t), V(t+h))(-h) - V(t)$$

for every $t \in R$ and $h \in D$.Therefore,

$$\tau_{-h}(\gamma, V(t+h)) - V(t) = \nabla(\gamma^\bullet(t), V(t+h))(-h) - V(t) = 0$$

for every $t \in R$ and $h \in D$. Now, change h to $-h$ and apply this equation to $t' = t+h$.

We notice the following consequence

Corollary 10 *A vector field* $V = (v^1, \ldots, v^n)$ *is parallel along* γ *iff it satisfies the following linear system of ordinary differential equations*

$$(*) \qquad dv^k/dt + \sum_{i,j=1}^{n} \Gamma_{ij}^k(\gamma(t))(d\gamma^i/dt)v^j(t) = 0$$

In particular, the well-known theorems on existence and uniqueness of solutions of (*) apply and we obtain

Corollary 11 *Given* $v \in M_{\gamma(0)}$ *there is a unique parallel vector field* V *along* γ *such that* $V(0) = v$.

This corollary allows us to transport a vector parallel to itself along a curve during any finite time. More precisely, we can extend the isomorphism $\tau_h(\gamma, -) : \pi^{-1}(\gamma(0)) \longrightarrow \pi^{-1}(\gamma(h))$ from $h \in D$ to an isomorphism $\tau_t(\gamma, -) : \pi^{-1}(\gamma(0)) \longrightarrow \pi^{-1}(\gamma(t))$ for any $t \in R$, by defining

$$\tau_t(\gamma, v) = V(t)$$

Still another consequence of (*) may be formulated as follows:

Corollary 12 *If* $\gamma : R \longrightarrow M$ *is a curve and* $v^1, \ldots, v^n$ *constitute an orthonormal basis of* $M_{\gamma(0}$, *there are unique parallel vector fields* $P^1, \ldots, P^n$ *along* γ *such that* $P^i(0) = v^i$ *and at each point* t, $\{P^1(t), \ldots, P^n(t)\}$ *is an orthonormal basis of* $M_{\gamma(t)}$.

Historical note: Parallel transport was introduced by Levi-Civita in 1917, after the publication of Einstein's fundamental work on General Relativity in 1916. He used (*) as definition and pointed out that the possibility of transporting a vector parallel to itself at a finite distance follows from these equations, a fact that he considers as "geometrically obvious".