General Relativity Fall 2018
Lecture 6: covariant derivatives

Yacine Ali-Haïmoud
(Dated: September 21, 2018)

Coordinate basis and dual basis – We saw that, given a coordinate system $\{x^\mu\}$, the partial derivatives $\partial_{(\mu)}$ are vector fields (defined in a neighborhood of p where the coordinates are defined), and moreover form a basis of the tangent space. Their actions on scalar functions are just $\partial_{(\mu)}|_p(f) \equiv \partial f/\partial x^\mu|_p$ (technically, it is $\partial/\partial x^\mu(f \circ \phi^{-1})$, where $\phi : \mathcal{M} \rightarrow \mathbb{R}^n$ is the chart or coordinate system).

We may define their dual basis $dx^{(\mu)}$, such that $dx^{(\mu)}(\partial_{(\nu)}) = \delta^\mu{}_\nu$. Note that I keep parentheses to signifiy that the (μ) is a label, rather than a component; moreover, $\partial_{(\mu)}$ is a *vector*, even if the label is downstairs, and $dx^{(\mu)}$ is a dual vector, even though the label is upstairs. The reason is that this works nicely with the sum-over-repeated-up-and-down-indices rule: a vector V^α can be decomposed as $V = V^\mu \partial_{(\mu)}$, and a dual vector W_α as $W = W_\mu dx^{(\mu)}$. When we write the metric tensor through the line element $ds^2 = g_{\mu\nu} dx^\mu dx^\nu$, we are in fact formally writing $g = g_{\mu\nu} dx^{(\mu)} \otimes dx^{(\nu)}$.

Gradient of a scalar function – Consider a smooth function $f : \mathcal{M} \rightarrow \mathbb{R}$. We define the dual vector field $(df)_\alpha \equiv (\nabla f)_\alpha \equiv \nabla_\alpha f$ as follows: for any vector field V^α, $(\nabla f)(V) \equiv V(f)$. Let us unpack this. A dual vector is a map from vectors to real number, so it makes sense to define the dual vector ∇f by its action on vectors. A vector is, by definition, a linear map from smooth functions to real numbers. Hence, at each point $p \in \mathcal{M}$, $V|_p(f)$ is a real number.

We can decompose ∇f on the coordinate dual basis $\{dx^{(\mu)}\}$: $\nabla f = \nabla_\mu f \; dx^{(\mu)}$. We now apply this to $\partial_{(\nu)}$. The left and right-hand side are given by

$$\nabla f(\partial_{(\nu)}) = \partial_{(\nu)}(f) = \frac{\partial f}{\partial x^\nu}, \tag{1}$$

$$\nabla_\mu f \; dx^{(\mu)}(\partial_{(\nu)}) = \nabla_\mu f \delta^\mu{}_\nu = \nabla_\nu f. \tag{2}$$

Therefore, the components of ∇f on the coordinate dual basis are just $\nabla_\mu f = \partial f/\partial x^\mu$. The dual vector field ∇f is called the gradient of f.

Covariant derivative of tensors: axiomatic definition – We now want to generalize the notion of a gradient to vectors and tensors. While it makes sense to compare a scalar function at different points (hence take its gradient), it is less obvious how one should do with vectors, which belong to different vector spaces at each point! The covariant derivative provides a geometric (i.e. coordinate independent) way to define the gradient of tensors.

A covariant derivative ∇ is a *linear* operator that, when acting on tensor fields of rank (k, l), gives back a tensor field of rank $(k, l+1)$. Given a tensor field $T^{\alpha_1..\alpha_k}{}_{\beta_1..\beta_l}$, we denote

$$(\nabla T)^{\alpha_1..\alpha_k}{}_{\beta_1..\beta_l\gamma} \equiv \nabla_\gamma T^{\alpha_1..\alpha_k}{}_{\beta_1..\beta_l} \equiv T^{\alpha_1..\alpha_k}{}_{\beta_1..\beta_l;\gamma}. \tag{3}$$

It must satisfy the following properties:
(i) When acting on functions (i.e. tensor fields of rank $(0,0)$), it just gives the gradient ∇f. So all covariant derivatives agree on scalar functions, and give the ordinary partial derivative in a coordinate basis.
(ii) It satisfies Leibniz' rule, i.e. $\nabla_\lambda (T^{\alpha ...}{}_{\beta ...} S^{\gamma ...}{}_{\delta ...}) = (\nabla_\lambda T^{\alpha ...}{}_{\beta ...}) S^{\gamma ...}{}_{\delta ...} + T^{\alpha ...}{}_{\beta ...} \nabla_\lambda S^{\gamma ...}{}_{\delta ...}$.
(iii) It commutes with contractions: given a tensor T^α_β, we define $\mathcal{T} \equiv T^\alpha_\alpha$, and require that $\nabla_\beta \mathcal{T} = \nabla_\beta T^\alpha_\alpha \equiv (\nabla T)^\alpha{}_{\alpha\beta}$.

These conditions do not fully determine the covariant derivative, as we will see later on.

Connection coefficients – Let us now pick some fixed coordinates $\{x^\mu\}$. We denote by $\partial^\alpha_{(\mu)}$ the associated coordinate basis vectors . The heavy notation is to emphasize that, for each label (μ), $\partial_{(\mu)}$ is a vecto field. Given a vector field $V^\alpha = V^\mu \partial^\alpha_{(\mu)}$, the coordinates V^μ in this basis can be seen as n scalar functions: they are just n real numbers. Again, here we have completely fixed the coordinate system and are not concerned with transformation of these numbers when we change the coordinates.

Take any covariant derivative ∇, and apply it to a vector field $V^\alpha = V^\mu \partial^\alpha_{(\mu)}$:

$$\begin{aligned} \nabla_\gamma V^\alpha &= \nabla_\gamma (V^\mu \partial^\alpha_{(\mu)}) = \nabla_\gamma (V^\mu) \partial^\alpha_{(\mu)} + V^\mu \nabla_\gamma \partial^\alpha_{(\mu)} \\ &= \frac{\partial V^\mu}{\partial x^\gamma} + \Gamma^\alpha{}_{\gamma\mu} V^\mu, \qquad \Gamma^\alpha{}_{\gamma\mu} \equiv \nabla_\gamma \partial^\alpha_{(\mu)}, \end{aligned} \tag{4}$$

where in the first line we used Leibniz' rule, and in the second line we used the fact that covariant derivatives, applied to scalar functions, are just the ordinary partial derivative. The object ${\Gamma^\alpha}_{\gamma\mu}$ is a *connection coefficient*: it connects the covariant derivative to the partial derivative in the chosen coordinate system.

The connection coefficient is *not a tensor*, and this can be seen as follows. Under a change of coordinates, we have

$$\frac{\partial V^{\mu'}}{\partial x^{\gamma'}} = \frac{\partial}{\partial x^{\gamma'}}\left(\frac{\partial x^{\mu'}}{\partial x^{\mu}} V^\mu\right) = \frac{\partial x^\gamma}{\partial x^{\gamma'}}\frac{\partial}{\partial x^\gamma}\left(\frac{\partial x^{\mu'}}{\partial x^{\mu}} V^\mu\right) = \frac{\partial x^\gamma}{\partial x^{\gamma'}}\frac{\partial x^{\mu'}}{\partial x^{\mu}}\frac{\partial V^\mu}{\partial x^\gamma} + \frac{\partial x^\gamma}{\partial x^{\gamma'}}\frac{\partial^2 x^{\mu'}}{\partial x^\mu \partial \gamma} V^\mu. \tag{5}$$

Because of the second piece, we see that $\frac{\partial V^{\mu'}}{\partial x^{\gamma'}}$ does not transform as a tensor. If $\nabla_\gamma V^\alpha$ is to be a tensor, then ${\Gamma^\alpha}_{\gamma\mu}$ cannot possibly transform as tensor. Said differently, the covariant derivative is made of two non-tensorial pieces: the partial derivative, and a "correction", proportional to the connection coefficient, which together, make it into a tensor! You will derive explicitly in homework 3 how the connection coefficient transforms under change of coordinates.

Covariant derivative of a dual vector field – Given Eq. (4), we can now compute the covariant derivative of a dual vector field W_α. To do so, pick an arbitrary vector field V^α, consider the covariant derivative of the scalar function $f \equiv V^\alpha W_\alpha$. This is the contraction of the tensor field ${T^\alpha}_\beta \equiv V^\alpha W_\beta$. Therefore, we have, on the one hand,

$$\nabla_\gamma (V^\alpha W_\alpha) = \nabla_\gamma f = \frac{\partial f}{\partial x^\gamma} = \frac{\partial}{\partial x^\gamma}(V^\alpha W_\alpha) = \frac{\partial V^\alpha}{\partial x^\gamma} W_\alpha + V^\alpha \frac{\partial W_\alpha}{\partial x^\gamma}, \tag{6}$$

since covariant derivatives applied to scalar functions just give the partial derivative. On the other hand, we have

$$\begin{aligned}\nabla_\gamma (V^\alpha W_\alpha) &= {(\nabla T)^\alpha}_{\alpha\gamma} \quad \text{[commutes with contractions]}\\ &= (\nabla_\gamma V^\alpha) W_\alpha + V^\alpha (\nabla_\gamma W_\alpha) \quad \text{[Leibniz rule]}\\ &= \left(\frac{\partial V^\alpha}{\partial x^\gamma} + {\Gamma^\alpha}_{\gamma\mu} V^\mu\right) W_\alpha + V^\alpha (\nabla_\gamma W_\alpha).\end{aligned} \tag{7}$$

Equating the two, we see that the term proportional to derivatives of V cancels out. Renaming dummy summed-over indices and factorizing V^α, we find that, for the equality to hold for any V^α, we must have

$$\nabla_\gamma W_\alpha = \frac{\partial W_\alpha}{\partial x^\gamma} - {\Gamma^\mu}_{\gamma\alpha} W_\mu. \tag{8}$$

More generally, for a tensor of arbitrary rank, the covariant derivative is the partial derivative plus a connection for each upper index, minus a connection for each lower index. You will derive this explicitly for a tensor of rank $(0, 2)$ in homework 3.

Torsion-free, metric-compatible covariant derivative – The three axioms we have introduced so far do not fully specify the covariant derivative: we can pick a coordinate system, and choose ${\Gamma^\alpha}_{\mu\nu} = 0$ in this coordinate system, and define a covariant derivative this way – it is just the partial derivative in that coordinate system, which obviously satisfies the 3 conditions of covaraint derivatives. The connection coefficient can then be obtained in any other coordinate system through its non-tensorial transformation law, and will not vanish in general. Said differently, we can define as many covariant derivatives as there are coordinate systems! So we need to be more specific to make progress.

We therefore add two conditions. First, we require that the covariant derivative is *torsion-free*, that is $\nabla_\alpha \nabla_\beta f = \nabla_\beta \nabla_\alpha f$, for any function f. It is easy to show, from Eq. (8) applied to $W_\alpha = \nabla_\alpha f$, that this implies that the connection coefficient is symmetric in its lower two indices.

Secondly, we impose *metric-compatibility*, i.e. that $\nabla_\alpha g_{\beta\gamma} = 0$. This ties the covaraint derivative to the metric tensor. You will derive in homework 3 that these two conditions uniquely characterize the connection coefficients: in any coordinate system, they are given by

$${\Gamma^\delta}_{\mu\nu} = \frac{1}{2} g^{\delta\lambda}\left(\frac{\partial g_{\lambda\mu}}{\partial x^\nu} + \frac{\partial g_{\lambda\nu}}{\partial x^\mu} - \frac{\partial g_{\mu\nu}}{\partial x^\lambda}\right). \tag{9}$$

This specific connection coefficients have a name, they are called the *Christoffel symbols*.