The Lie Derivative

Charles Daly

Summary

These notes are dedicated to some thoughts I've had on the Lie derivative. A while back when I first saw the definition of the Lie derivative and understood nearly nothing about it, William Goldman told me one of those facts that just stuck with me. Specifically that the Lie derivative of vector fields X and Y may be thought of as the double derivative of the commutator of their respective flows. This seems to be one of those well known facts of differential geometry, but I haven't quite found a resource that has the proof in full, so I thought I would write it up just in case anyone else out there was just as hopelessly lost as I was when I first saw the Lie derivative.

1 Motivation

First and foremost, everything which follows is assumed to smooth unless otherwise stated. Let M and N be a smooth manifolds of dimension m and n respectively. Given a smooth map $F : M \longrightarrow N$, we may associate to F a bundle homomorphism $dF : TM \longrightarrow TN$. This map is sometimes referred to as the *global differential of* F. It is the manifold analogue of the standard differential of a smooth map $G : \mathbb{R}^m \longrightarrow \mathbb{R}^n$, and consequently may be thought of as the derivative of F. The global differential makes its appearance almost immediately in any introductory differential geometry course.

When I first saw the global differential it kind of felt like a power-up in a video game. Suddenly I could start taking derivatives of any maps I wanted and make perfectly good sense of them. While to some extent this is true, I remember spending hours upon hours trying to reconcile the following question. *What is the derivative of a vector field?* This seemingly innocuous question leads to a wealth of very interesting mathematics including the notion of *connections on manifolds* and *Ehresmann connections*.

Surely given a smooth vector field Y on M, it enjoys the property of being a smooth map $Y : M \longrightarrow TM$ so that $\pi \circ Y = \text{id}$ where $\pi : TM \longrightarrow M$ is the standard tangent bundle projection. Consequently the derivative of a vector field makes perfectly good sense, and its global differential is given by $dY : TM \longrightarrow T(TM)$. And perhaps now you can see some technical difficulties here. In particular, if we let $p \in M$, then $dY_p : T_pM \longrightarrow T_{Y_p}(TM)$ is a linear map from the tangent space of the *manifold* M at p to the tangent space of the *tangent bundle of* M at Y_p. While this is all theoretically correct, for each $v \in T_pM$, we would like $dY_p(v)$ to be an element of T_pM as well. This way for we could for example, make some comparisons between v and $dY_p(v)$.

I think it's well worth exploring several potential 'work arounds' before simply giving up and calling it a day. For example, perhaps one could take the vector $dY_p(v) \in T_{Y_p}(TM)$ and identify it with its image under $d\pi_{Y_p} : T_{Y_p}(TM) \longrightarrow T_pM$. This is after all the space we would like $dY_p(v)$ to have been in anyways. But note that by the chain rule

$$d\pi_{Y_p}\left(dY_p(v)\right) = d\left(\pi \circ Y\right)_p(v) = d\left(\text{id}\right)_p(v) = v \tag{1}$$

We can see that in our attempt to identify $dY_p(v)$ with a vector in the tangent space via the projection map, we've essentially disregarded all the data about Y itself. This shouldn't come as too much as a surprise though. If you think of derivatives in terms of directions and curves for example, we may take a smooth path $\gamma : I \longrightarrow M$ representing v, wherein $\gamma(0) = p$ and $\gamma'(0) = v$. In so doing, $dY_p(v)$ is represented by $(Y \circ \gamma)'(0)$. Note that this may be thought of as taking the time derivative of the vector field along γ given by $Y_{\gamma(t)}$. I think this perspective provides some insight into the difficulties here. Specifically, $Y_{\gamma(t)}$ may be thought of as a lift of γ to the tangent bundle. In terms of commutative diagrams,

$$\begin{array}{ccc} & & TM \\ & \overset{Y_\gamma}{\nearrow} & \big\downarrow \pi \\ I & \xrightarrow[\gamma]{} & M \end{array} \tag{2}$$

Now note that as t varies, $Y_{\gamma(t)} \in T_{\gamma(t)}M$ for each $t \in I$. That said, if we think about TM as a bundle over M, we can see the difficulty in differentiation. The curve $Y_{\gamma(t)}$ is changing fibres of $\pi : TM \longrightarrow M$ as t varies which translates to $Y_{\gamma(t)}$ lying in different tangent spaces for different t. Ideally, one would like to try and define

$$dY_p(v) := \lim_{t \to 0} \frac{1}{t} \left(Y_{\gamma(t)} - Y_p \right) \tag{3}$$

But notice, that $Y_{\gamma(t)}$ and Y_p lie in different tangent spaces, and thus it makes no sense to subtract them. Figure 1 below illustrates this difficulty. Moreover, when we compose $Y_{\gamma(t)}$ with the projection map, we simply end up with $\pi \left(Y_{\gamma(t)} \right) = \gamma(t)$, thus eliminating all information pertaining the vector field in question. The vector field Y tells us what vectors to assign each value of t, $Y_{\gamma(t)}$, whereas the curve γ corresponding to v tells us which direction along the manifold we're evaluating Y in. This sort of mode of thought helps lead to the notion of an *Ehresmann connection*, which is in essence a way to split $T(TM)$ into two bundles, one a vertical bundle corresponding the kernel of the map $d\pi : T(TM) \longrightarrow TM$, and a horizontal bundle complementary to the vertical bundle. The vertical bundle is well defined independent of the connection, whereas the horizontal bundle is a choice that in essence allows us to decompose $dY_p(v)$ into a vertical and horizontal component. More details about this can be found in some excellent notes by Timothy Goldberg, *What is a Connection, and What is is Good For?* [1]

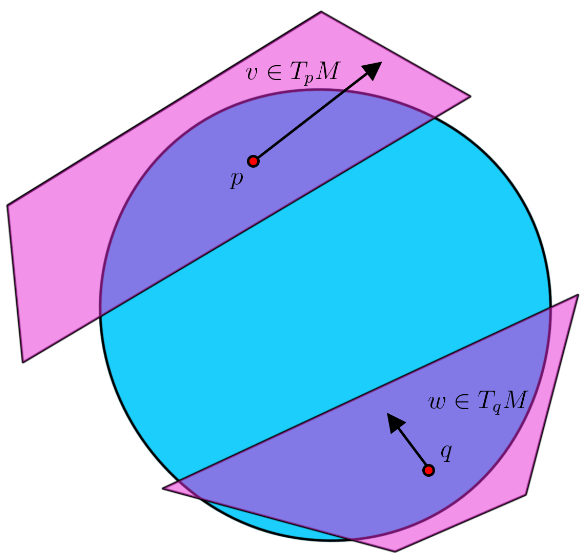

Figure 1: Note here on the sphere $M = S^2$ there are two points $p, q \in M$ and their tangent spaces T_pM and T_qM are illustrated in pink. The vectors v and w cannot be naturally subtracted as they live in different tangent spaces. This difficulty arises because there's no natural isomorphism between T_pM and T_qM.

To conclude this section, I would like to point out there are situations where these difficulties do not arise. Specifically, in the case of $\mathbb{R}^n$, we have the tangent bundle naturally identifies with $\mathbb{R}^n \times \mathbb{R}^n$. The first factor may be thought of as the manifold component, whereas the second factor may be thought of as the tangent space component. In this situation, a vector field Y on $\mathbb{R}^n$ is a map $Y : \mathbb{R}^n \longrightarrow T(\mathbb{R}^n) \simeq \mathbb{R}^n \times \mathbb{R}^n$ such that $\pi_1 \circ Y = \mathrm{id}$ where $\pi_1 : \mathbb{R}^n \times \mathbb{R}^n \longrightarrow \mathbb{R}^n$ is projection onto the first factor. Consequently the first factor of Y is the identify, and that said, we may as well identify Y as map $Y : \mathbb{R}^n \longrightarrow \mathbb{R}^n$, corresponding to the second factor.

Now to each tangent vector $v \in T_p\mathbb{R}^n \simeq \mathbb{R}^n$, we may take the derivative of Y in the direction of v. Motivated by Equation 3, let us represent v by the curve $\gamma(t) = (\gamma^1(t), \dots, \gamma^n(t))$ so that $\gamma(0) = p$ and $\gamma'(0) = v$. Similarly, let us express Y in coordinates by $Y(q) = (Y^1(q), \dots, Y^n(q))$. Returning to equation 3 we have

$$\begin{aligned} dY_p(v) &= \lim_{t\to 0} \frac{1}{t}\left(Y_{\gamma(t)} - Y_p\right) = \lim_{t\to 0}\frac{1}{t}\left(Y^i(\gamma(t)) - Y^i(p)\right)e_i = \frac{\partial Y^i}{\partial x^j}(\gamma(0))\frac{d\gamma^j}{dt}(0)e_i \\ &= \left(\frac{\partial Y^1}{\partial x^j}(p)\frac{d\gamma^j}{dt}(0), \dots, \frac{\partial Y^n}{\partial x^j}(p)\frac{d\gamma^j}{dt}(0)\right) \end{aligned} \tag{4}$$

where the third equation in the top row follows by the chain rule applied to the composition $Y \circ \gamma$. Note that Equation 4 provides us with a notion of a derivative of a vector field in $\mathbb{R}^n$, and moreover, it's kind of what anyone would guess, namely just differentiate the components. At face value, this really just appears to be an exercise in coordinate expansions. In fact, it's worth trying to see how this translates over in the general case of a manifold.

For an arbitrary manifold, we no longer enjoy the ability to have a global trivialization of the tangent bundle, i.e. TM is not necessarily identified with $M \times \mathbb{R}^n$. Nevertheless, we still enjoy local trivializations of the tangent bundle. So that said, for a point $p \in M$, pick a chart (U, ϕ) about it where $\phi : U \longrightarrow \mathbb{R}^n$ is a diffeomorphism onto some open subset of $\mathbb{R}^n$. This induces a local frame for the tangent bundle given by $\{\partial/\partial x^i\}$, so we may express a vector field, at least locally about p, as $Y|_q = Y^i(q)\partial/\partial x^i|_q$ for all $q \in U$. In addition if we pick a $v \in T_pM$, we may express it in terms of this frame as $v = v^j\partial/\partial x^j|_p$. That said, we may try and define

$$dY_p(v) := v^j \frac{\partial Y^i}{\partial x^j}(p)\frac{\partial}{\partial x^i}\bigg|_p \tag{5}$$

just as we did in Equation 4 where each v^j corresponds to $d\gamma^j/dt(0)$. As with most constructions on manifolds, we would ideally like to ensure this quantity is well defined independent of coordinate patch. To this end, if (V, ψ) is another coordinate patch, with coordinates $\tilde{x}^k$, we note that have the equations $Y|_q = Y^i(q)\partial/\partial x^i|_q = \widetilde{Y}^k(q)\partial/\partial\tilde{x}^k|_q$ and $v = v^j\partial/\partial x^j|_p = \tilde{v}^l\partial/\partial\tilde{x}^l|_p$. Their coordinates are related by

$$Y^i(q) = \widetilde{Y}^k(q)\frac{\partial x^i}{\partial\tilde{x}^k}(q) \text{ and } v^j = \tilde{v}^l\frac{\partial x^j}{\partial\tilde{x}^l}(p) \tag{6}$$

Differentiating the first equation for $Y^i(q)$ and suppressing the evaluation at at p yields

$$\frac{\partial Y^i}{\partial x^j} = \frac{\partial}{\partial x^j}\left(\widetilde{Y}^k\frac{\partial x^i}{\partial\tilde{x}^k}\right) = \frac{\partial\tilde{x}^m}{\partial x^j}\frac{\partial}{\partial\tilde{x}^m}\left(\widetilde{Y}^k\frac{\partial x^i}{\partial\tilde{x}^k}\right) = \frac{\partial\tilde{x}^m}{\partial x^j}\frac{\partial\widetilde{Y}^k}{\partial\tilde{x}^m}\frac{\partial x^i}{\partial\tilde{x}^k} + \widetilde{Y}^k\frac{\partial\tilde{x}^m}{\partial x^j}\frac{\partial^2 x^i}{\partial\tilde{x}^m\partial\tilde{x}^k} \tag{7}$$

Substituting Equation 7 back into Equation 5 with the change of coordinate for v^j in Equation 6 yields

$$\begin{aligned}
dY_p(v) &= \tilde{v}^l \frac{\partial x^j}{\partial \tilde{x}^l}\left(\frac{\partial \tilde{x}^m}{\partial x^j}\frac{\partial \widetilde{Y}^k}{\partial \tilde{x}^m}\frac{\partial x^i}{\partial \tilde{x}^k} + \widetilde{Y}^k \frac{\partial \tilde{x}^m}{\partial x^j}\frac{\partial^2 x^i}{\partial \tilde{x}^m \partial \tilde{x}^k}\right)\frac{\partial}{\partial x^i} \\
&= \tilde{v}^l \left(\frac{\partial x^j}{\partial \tilde{x}^l}\frac{\partial \tilde{x}^m}{\partial x^j}\right)\frac{\partial \widetilde{Y}^k}{\partial \tilde{x}^m}\left(\frac{\partial x^i}{\partial \tilde{x}^k}\frac{\partial}{\partial x^i}\right) + \tilde{v}^l \widetilde{Y}^k \left(\frac{\partial x^j}{\partial \tilde{x}^l}\frac{\partial \tilde{x}^m}{\partial x^j}\right)\frac{\partial^2 x^i}{\partial \tilde{x}^m \partial \tilde{x}^k}\frac{\partial}{\partial x^i} \\
&= \tilde{v}^l \delta_l^m \frac{\partial \widetilde{Y}^k}{\partial \tilde{x}^m}\frac{\partial}{\partial \tilde{x}^k} + \tilde{v}^l \widetilde{Y}^k \delta_l^m \frac{\partial^2 x^i}{\partial \tilde{x}^m \partial \tilde{x}^k}\frac{\partial}{\partial x^i} \\
&= \tilde{v}^l \frac{\partial \widetilde{Y}^k}{\partial \tilde{x}^l}\frac{\partial}{\partial \tilde{x}^k} + \tilde{v}^l \widetilde{Y}^k \frac{\partial^2 x^i}{\partial \tilde{x}^l \partial \tilde{x}^k}\frac{\partial}{\partial x^i} \qquad (8)
\end{aligned}$$

While these calculations are quite lengthy, they do provide some insight into what goes wrong in our definition. If we compare Equation 8 with Equation 5, we see precisely what goes wrong. In particular, the presence of this second term in Equation 8 is an obstruction to this quantity being well defined independent of coordinate chart.

That said, we should take the time to consider why this works in the case of $\mathbb{R}^n$ and not a general manifold. Note that in Equation 8, if we hypothetically had our manifold M covered with a coordinate atlas $\{(U_\alpha, \phi_\alpha)\}$ such that the change of coordinates had vanishing Hessian, then Equation 5 is well defined, and this derivative is dependent on the coordinate atlas. In fact a very natural class of manifolds to look at for which this works are *affine manifolds*. Affine manifolds are a class of manifolds whose change of coordinates are locally affine maps, namely maps from $\mathbb{R}^n \longrightarrow \mathbb{R}^n$ of the form $F(x) = Ax + b$, where $A \in \mathrm{GL}(n, \mathbb{R})$ and $b \in \mathbb{R}^n$. For those of you familiar with connections, it's worth verifying this derivative corresponds to the pullback connection of the standard connection on $\mathbb{R}^n$.

Another case where we may define the derivative of a vector field without necessarily deferring to the machinery of a connection is in the case where M is parallelizable, that is, there's a bundle isomorphism between TM and $M \times \mathbb{R}^n$. As per consequence a vector field on M as a smooth map $Y : M \longrightarrow TM$ is the same a having a smooth map $Y : M \longrightarrow M \times \mathbb{R}^n$ where Y is the identity on the first component. Consequently, the data of a vector field is really just a map $Y : M \longrightarrow \mathbb{R}^n$. We may now define $dY_p(v)$ to be as in Equation 3, where the difference $Y_{\gamma(t)} - Y_p$ makes sense now as per consequence of the global trivialization of the tangent bundle. Certainly we may always trivialize locally, as we attempted to in Equation 8, but it seems unreasonable to have this difference quotient depend on local identifications, whereas it's far more palatable for it to depend on a single fixed identification. That said, it is perhaps again worth emphasizing that such a definition is indeed still relying on a choice, in particular, here we are relying on an identification between TM and $M \times \mathbb{R}^n$ and in general this identification is non-canonical.

2 Lie Derivative

Given the above discussion, it may seem hopeless to define a meaningful derivative of a vector field on a manifold. One of the main issues is the inability to compare tangent spaces. This is a fundamental difficulty of differential geometry and has been studied extensively. In fact, one will often find calculus problems about *rolling without slipping* which are intimately related to notions of identifying tangents spaces at different points via a construction known as *parallel transport*. The idea is parallel transport provides us a means of identifying tangent spaces at different points via curves on the manifold. Certainly how one travels along the manifold, will in general, impact these identifications. Take for example the two curves on the sphere and in Figure 2 below.

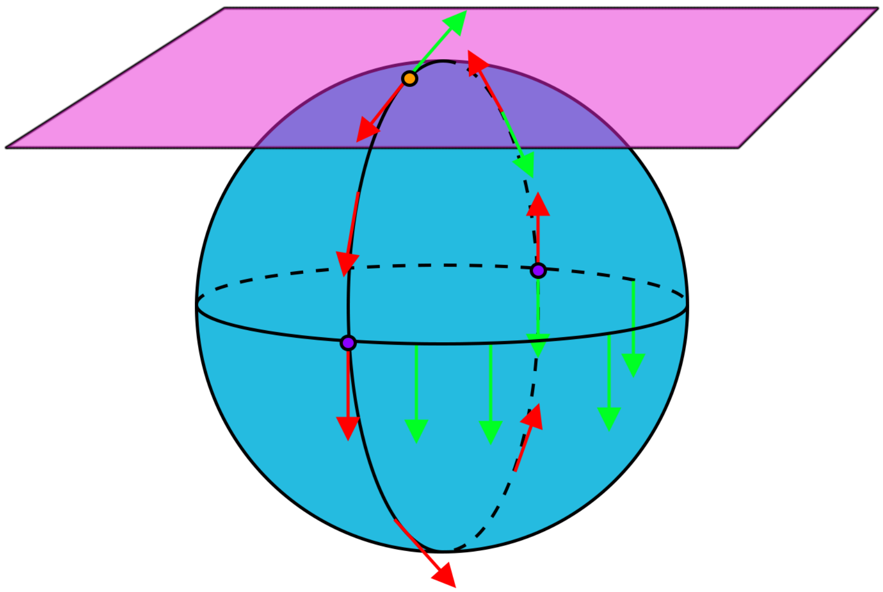

Figure 2: Here the vector v in red located at the north pole $p \in M$ in orange is parallel transported along two curves. The first is the vertical circle or prime meridian. Note that upon returning the point p, the vector v returns to its original position. Contrast this with the situation where halfway down when the curve hits the first purple point, then starts to transport the vector around the equator which is labeled in green. After going around half the equator until the second purple point, the vector is then transported back along the prime meridian. The resulting vector in green returns original tangent space, but is in fact equal to the $-v$. This is evidence that the sphere is not flat.

Motivated by this idea of looking at identifications between tangent spaces, one is perhaps led to consider the diffeomorphism group of a manifold. If we want to make sense of the quantity $Y_{\gamma(t)} - Y_p$, we need some way of identifying tangent spaces. Diffeomorphisms certainly provide this. In particular, if F is a diffeomorphism of M taking a point p to another point q, then the differential $dF_p : T_pM \longrightarrow T_qM$ is an isomorphism between vector spaces. Now let's say we have a path of diffeomorphisms, F_t, such that $F_t(p) = \gamma(t)$. This provides us with identifications of T_pM and $T_{\gamma(t)}M$ via $d(F_t)_p$. Under the circumstances we can't necessarily make sense of $Y_{\gamma(t)} - Y_p$, but it seems then natural to consider the quantity

$$(d(F_t)_p)^{-1}\left(Y_{\gamma(t)}\right) - Y_p \tag{9}$$

Before proceeding, let's make sure this is a reasonable quantity. Certainly $Y_{\gamma(t)} \in T_{\gamma(t)}M$. On the other hand, since $d(F_t)_p : T_pM \longrightarrow T_{\gamma(t)}M$, we know the inverse map, which is well defined as F_t is a diffeomorphism, takes $T_{\gamma(t)}M$ to T_pM. Thus, the quantity in equation 9 does indeed make sense. Moreover, it is feasible that we could then take the limit

$$\lim_{t\to 0}\frac{1}{t}\left((d(F_t)_p)^{-1}\left(Y_{\gamma(t)}\right) - Y_p\right) \tag{10}$$

For a general path of diffeomorphisms, one should not expect the above to converge, as we have no control over what $d(F_t)_p$ does at $t = 0$ for example, it's merely a map that fixed p under our current stipulations. In addition, how does one obtain a path of diffeomorphisms $(F_t)_p = \gamma(t)$ where $\gamma(t)$ represents a vector $v \in T_pM$?

A very natural context to take into consideration is flows on a manifold. Given a one parameter subgroup of a diffeomorphisms $F_t : \mathbb{R} \longrightarrow \mathrm{Diff}(M)$, one knows that $F_0 = \mathrm{id}$, and thus the behavior at $t = 0$, and ideally locally about $t = 0$, must be reasonably controlled.

Given the above observations, it's not entirely outlandish then then try and differentiate a vector field with respect to another vector field as vector fields define local flows. That said, let X and Y be vector fields on a manifold M. For the sake of simplicity, I'm just going to assume their corresponding flows are indeed complete, though the following constructions and arguments may be adapted to purely local arguments.

As Y is the vector field we wish to differentiate, X is the vector field we intend to differentiate along, let $\theta : \mathbb{R} \times M \longrightarrow M$ be the corresponding flow map of X. Let $p \in M$, and we wish to define the derivative of Y with respect to X as p, in a similar fashion to Equation 10. Denoting this derivative $\mathcal{L}_X Y|_p$ we define

$$\begin{aligned}\mathcal{L}_X Y|_p &:= \lim_{t\to 0}\frac{1}{t}\left((d(\theta_t)_p)^{-1}\left(Y_{\theta_t(p)}\right) - Y_p\right)\\ &= \lim_{t\to 0}\frac{1}{t}\left(d(\theta_{-t})_{\theta_t(p)}\left(Y_{\theta_t(p)}\right) - Y_p\right)\end{aligned} \tag{11}$$

where the second equality follows from the fact that $\theta_t^{-1} = \theta_{-t}$. Several facts need to be verified and will not be provided in these notes, the most pressing being that this vector field is indeed well defined. For those interested in further detail, and where much of the notation and inspiration for these notes has been drawn from, I refer you to Jack Lee's outstanding differential geometry text *Introduction to Smooth Manifolds* [2]. This book is a really solid general reference for the material, and his conversational writing style, thoroughly detailed arguments, and illustrations make this one of those texts I just roundly recommend to everyone.

Before continuing it worth pointing out the definition as in Equation 11 is very different than the one provided in Equation 5. Recall the method of identification relating $T_{\gamma(t)}\mathbb{R}^n$ and $T_p\mathbb{R}^n$ was via the identification induced by translation taking p to $\gamma(t)$. Here though the identification between $T_{\gamma(t)}\mathbb{R}^n$ and $T_p\mathbb{R}^n$ is induced by the flow of a vector field. The following examples serve to illustrate this difference.

Example 1. Let $M = \mathbb{R}^2$ and define

$$X = -y\frac{\partial}{\partial x} + x\frac{\partial}{\partial y} \text{ and } Y = \frac{\partial}{\partial x} \tag{12}$$

Let $p = (x, y)$ and let us calculate $\mathcal{L}_X Y$. Deferring to Equation 11 we first calculate the flow of X. In fact, one can show with relative ease that the corresponding system of differential equations is given by

$$x' = -y \text{ and } y' = x \text{ where } (x(0), y(0)) = (x, y) \implies \theta_t(x, y) = \begin{pmatrix} \cos(t) & -\sin(t) \\ \sin(t) & \cos(t) \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

That said, $Y_{\theta_t(p)} = \partial/\partial x|_{\theta_t(p)}$ and so

$$d\left(\theta_{-t}\right)_{\theta_t(p)}\left(Y_{\theta_t(p)}\right) = d\left(\theta_{-t}\right)_{\theta_t(p)}\left(\left.\frac{\partial}{\partial x}\right|_{\theta_t(p)}\right) = \cos(t)\left.\frac{\partial}{\partial x}\right|_p - \sin(t)\left.\frac{\partial}{\partial y}\right|_p$$

The minus sign is because we are using the inverse flow, θ_{-t}, and not θ_t. Plugging this result into Equation 11 yields

$$\mathcal{L}_X Y = \lim_{t\to 0}\frac{1}{t}\left(\cos(t)\left.\frac{\partial}{\partial x}\right|_p - \sin(t)\left.\frac{\partial}{\partial y}\right|_p - \left.\frac{\partial}{\partial x}\right|_p\right) = \lim_{t\to 0}\frac{\cos(t)-1}{t}\left.\frac{\partial}{\partial x}\right|_p - \frac{\sin(t)}{t}\left.\frac{\partial}{\partial y}\right|_p = -\left.\frac{\partial}{\partial y}\right|_p$$

Again it's all too easy to get lost in the abstract calculations. This example is intriguing for the following reason. Generally speaking we think of the vector field $\partial/\partial x$ as 'constant.' It's invariant under the standard identification of $T_p\mathbb{R}^2$ and $T_q\mathbb{R}^2$ induced by the translation taking p to q. This though is not how we are identifying tangent spaces. Here we are locally identifying tangent spaces in the same orbits of the flow of X via flow out map, namely θ_t. In our case these are circles. Now as a vector travels along a circle, it seems reasonable that it indeed may change its direction, and this is precisely what's happening and why we're getting a non-trivial Lie derivative. Figure 3 illustrates this process.

Example 2. Let $M = \mathbb{R}^2$ and define

$$X = -y\frac{\partial}{\partial x} + x\frac{\partial}{\partial y} \text{ and } Y = x\frac{\partial}{\partial x} + y\frac{\partial}{\partial y} \tag{13}$$

In a similar fashion to the example above, one may do the calculations to figure out that Y is invariant under the flow of X and thus the Lie derivative in this case is indeed 0. It is worth verifying these calculations, but a picture is provided below to illustrate this.

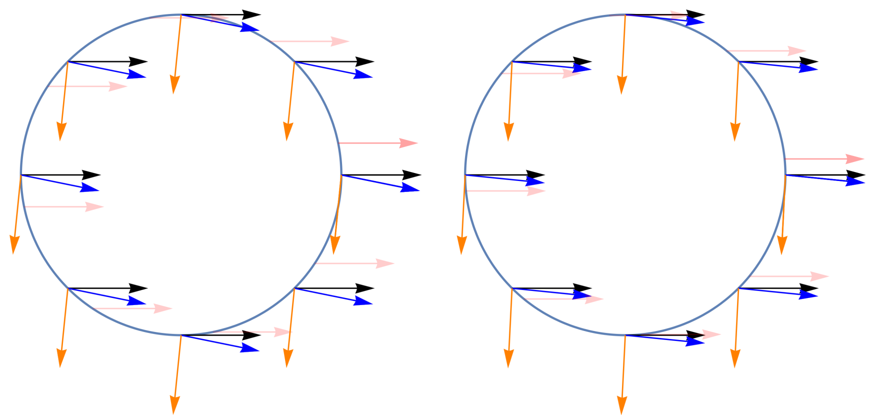

Figure 3: Here the vector field $Y = \partial/\partial x$ is illustrated in black. The flow of X at a point is depicted by the circle in light blue. The vector field $Y_{\theta_t(p)}$ is illustrated in light red. Its pull back under the flow of X is illustrated in blue, which note, is different than Y. The difference quotient is illustrated in orange. The left image is for $t = 0.2$ and the right is $t = 0.1$.

Example 3. Let $M = \mathbb{R}^2$ and define

$$X = (x+y)\frac{\partial}{\partial x} + y\frac{\partial}{\partial y} \text{ and } Y = x\frac{\partial}{\partial y} \tag{14}$$

Here is an example where the vector field Y vanishes along the line $x = 0$, but its derivative along the Y axis is generally non-zero. Again, without too much difficulty one can verify that the flow of X is given by

$$\theta_t(x,y) = \begin{pmatrix} e^t & te^t \\ 0 & e^t \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \implies Y_{\theta_t(p)} = \left(e^t x + te^t y\right) \frac{\partial}{\partial y}\bigg|_{\theta_t(p)}$$

Calculating the limit of the difference quotient yields

$$\begin{aligned} \mathcal{L}_X Y &= \lim_{t\to 0} \frac{1}{t}\left(\left(e^t x + te^t y\right)\left(-te^{-t}\frac{\partial}{\partial x}\bigg|_p + e^{-t}\frac{\partial}{\partial y}\bigg|_p\right) - x\frac{\partial}{\partial y}\bigg|_p\right) \\ &= \lim_{t\to 0} -(x+ty)\frac{\partial}{\partial x}\bigg|_p + ye^{-t}\frac{\partial}{\partial y}\bigg|_p = -x\frac{\partial}{\partial x}\bigg|_p + y\frac{\partial}{\partial y}\bigg|_p \end{aligned}$$

which note, indeed does not vanish on the line $x = 0$ except at the origin. Figure 5 below illustrates the calculation along two different integral curves of X.

It should be mentioned there is yet another formulation of the Lie derivative in terms of vector fields acting on the algebra of smooth real-valued functions on a manifold M. Specifically, given two vectors X and Y on a manifold M, one may define the Lie-bracket of X and Y to be the vector field which acts on the algebra of smooth real-valued functions on M via $(XY - YX)f = X(Yf) - Y(Xf)$ for all $f \in \mathrm{C}^\infty(M)$.

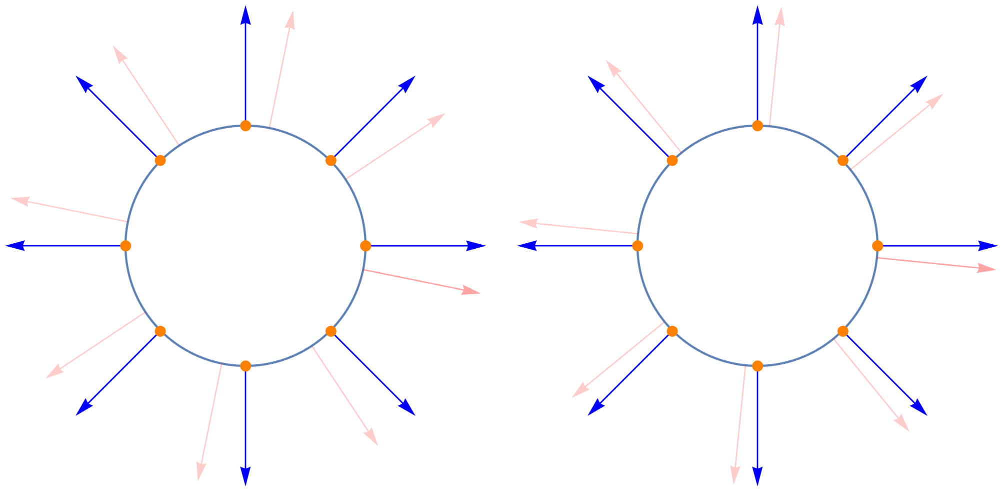

Figure 4: Here the vector field $Y_{\theta_t(p)}$ is illustrated in red and its pullback, under the flow of X, is illustrated in blue. The original vector field Y is illustrated in black, but is obscured by the pullback in blue, as they are equal. The difference quotient is illustrated by the orange zero vectors. The left image is for $t = 0.2$ and the right is $t = 0.1$.

While the following formulation of the Lie derivative is substantially more computationally accessible, and theoretically advantageous to work with, it suffers the serious draw back of not enjoying any 'obvious' geometric interpretation, at least to my knowledge. In fact, that it the above even defines a vector field is far from obvious in my opinion. Nevertheless, it allows one to calculate Lie-derivatives with relative ease and has a straightforward coordinate representation. If X and Y are locally given in coordinates by $X = X^i \partial/\partial x^i$ and $Y = Y^j \partial/\partial x^j$, then the Lie-bracket of these two vector fields is given by

$$\mathcal{L}_X Y = [X, Y] = \left(X^i \frac{\partial Y^j}{\partial x^i} - Y^i \frac{\partial X^j}{\partial x^i} \right) \frac{\partial}{\partial x^j} \tag{15}$$

That this definition and the formulation in terms of derivatives of flows agree is far from clear, but nevertheless one may find a proof of this in Jack Lee's text [2].

3 Derivative of Commutator

As promised in the summary, we conclude by providing a proof that the Lie derivative of vector fields X and Y on M is given by the second derivative of the commutator of their flows. Let X and Y have flows given by θ and ϕ. Pick a point $p \in M$, and a coordinate patch centered about p, $(V, \psi) : V \longrightarrow \mathbb{R}^n$. Assume this coordinate patch is sufficiently small so that there's a box of $(-\epsilon, \epsilon)^2$ for which both θ and ϕ are defined everywhere in V about p. This stipulation is to guarantee that the flows do not blow up in finite time, and instead become undefined only in so far as they leave the open set V. As this discussion is purely local, we may assume in fact identify V with an open subset of $\mathbb{R}^n$ centered at the origin.

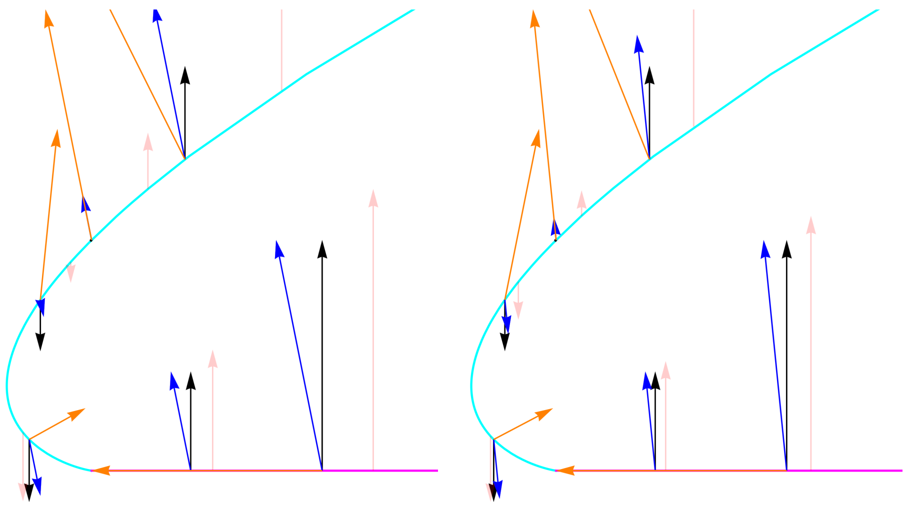

Figure 5: The vector field $Y = x\partial/\partial y$ is depicted in black along two integral curves of the vector field $X = (x+y)\partial/\partial x + y\partial/\partial y$ one in cyan and another in magenta. The vector field in light red is Y evaluated at $\theta_t(p)$ and its pullback under the flow of X is depicted in blue. The difference quotient is depicted in orange. Note there's a place where the orange and blue vectors are scalar multiples of each other. This point corresponds to a point where Y vanishes, yet its Lie derivative is non-zero. The left image is for $t = 0.2$ and the right is $t = 0.1$.

Consider the map $(\phi_{-t}\theta_{-t}\phi_t\theta_t)(0) : (-\epsilon, \epsilon) \longrightarrow \mathbb{R}^n$. If we Taylor expand both $\theta : (-\epsilon, \epsilon) \times V \longrightarrow V$ and $\phi_t : (-\epsilon, \epsilon) \times V \longrightarrow V$ about the origin, we have the following

$$\begin{aligned}\theta(t,x) &= \theta(0,0) + \frac{\partial\theta}{\partial t}(0,0)t + \frac{\partial\theta}{\partial x}(0,0)x + \frac{1}{2}\bigg(\frac{\partial^2\theta}{\partial t^2}(0,0)t^2 \\ &\qquad + \frac{\partial^2\theta}{\partial t\partial x}(0,0)tx + \frac{\partial^2\theta}{\partial x\partial t}(0,0)tx + x^T\frac{\partial^2\theta}{\partial x^2}(0,0)x\bigg) + \dots \\ &= 0 + \frac{\partial\theta}{\partial t}(0,0)t + \frac{\partial\theta}{\partial x}(0,0)x + \frac{1}{2}\frac{\partial^2\theta}{\partial t^2}(0,0)t^2 + \frac{\partial^2\theta}{\partial x\partial t}(0,0)tx + \frac{1}{2}x^T\frac{\partial^2\theta}{\partial x^2}(0,0)x + \dots \end{aligned} \tag{16}$$

where here I'm being very abusive in notation. In particular, the expression $(\partial^2\theta/\partial x^2)(0,0)$ is supposed to denote the second derivative of a vector valued function at the point $(0,0)$. If θ took values in $\mathbb{R}$, you get the standard Taylor expansion of a map from $\mathbb{R}^m \longrightarrow \mathbb{R}$ and this quantity represents the Hessian matrix. In our case though, this is a tensor quantity, which nevertheless can be recovered by doing Taylor expansions on each θ^i in the coordinate expansions of $\theta = (\theta^1, \theta^2, \dots, \theta^n)$. Unfortunately I don't know if there's a standard notation for this, so I just kind of made up my own, but if you need any convincing of what this is, just consider the two by two case from $\mathbb{R}^2$ to $\mathbb{R}^2$ and write out the details to see exactly what it is.

We may calculate the derivatives in Equation 16, as they reduce quite a bit. Recall the first input of θ is the time coordinate, whereas the second input is the spatial coordinate.

We have the following equalities

$$\frac{\partial\theta}{\partial t}(0,0) = \frac{\partial}{\partial t}\bigg|_{t=0} \theta(t,0) = X(0) \tag{17}$$

$$\frac{\partial\theta}{\partial x}(0,0) = \frac{\partial}{\partial x}\bigg|_{x=0} \theta(0,x) = \mathrm{id} \tag{18}$$

$$\frac{\partial^2\theta}{\partial t^2}(0,0) = \frac{\partial^2}{\partial t^2}\bigg|_{t=0} \theta(t,0) = \frac{\partial}{\partial t}\bigg|_{t=0} X_{\theta_t(0)} = \frac{\partial X}{\partial x}(0)X(0) \tag{19}$$

$$\frac{\partial^2\theta}{\partial x \partial t}(0,0) = \frac{\partial}{\partial x}\bigg|_{x=0} \frac{\partial}{\partial t}\bigg|_{t=0} \theta(t,x) = \frac{\partial}{\partial x}\bigg|_{x=0} X(x) = \frac{\partial X}{\partial x}(0) \tag{20}$$

$$\frac{\partial^2\theta}{\partial x^2}(0,0) = \frac{\partial^2}{\partial x^2}\bigg|_{x=0} \theta(0,x) = \frac{\partial^2}{\partial x^2}\bigg|_{x=0} x = 0 \tag{21}$$

Line 17 follow immediately from the relation between vector fields and the induced flow, namely the flow differentiates in time to the vector field. Line 18 is a consequence of the fact that $\theta(0,x)$ is the identity map on the space, so its derivative is the identity map on tangent spaces. Line 19 is again a consequence of the properties of flows, but where the last equality follows from the chain rule applied to $(X \circ \theta)(t,0)$. Line 20 is straight forward and follows more or less by definition. Line 21 follows again by consequence of the fact that $\theta(0,x)$ is the identity map on the space, so its second derivative vanishes. To be convinced of 21 it helps to think of the array of second derivatives of the function $(x_1,\dots,x_n) \mapsto (x_1,\dots,x_n)$. The first derivative may be thought of as the identity matrix, and differentiating the components of a matrix with constant values should yield zero. Combining the results of Lines (17)-(21) provides the second order Taylor expansion of θ at $(0,0)$ below.

$$\theta(t,x) = x + t\left(X(0) + \frac{\partial X}{\partial x}(0)x\right) + \frac{t^2}{2}\frac{\partial X}{\partial x}(0)X(0) + \dots \tag{22}$$

The same holds true for ϕ,

$$\phi(t,y) = y + t\left(Y(0) + \frac{\partial Y}{\partial x}(0)y\right) + \frac{t^2}{2}\frac{\partial Y}{\partial x}(0)X(0) + \dots \tag{23}$$

Composing Equations 22 and 23 we may calculate the commutator of the flows at $0 \in V$. As these calculations are pretty intense, I used Mathematica, see Figure 6 below for the exact code. In particular, this yields

$$\phi_{-t}\theta_{-t}\phi_t\theta_t(0) = t^2\left(\frac{\partial Y}{\partial x}(0)X(0) - \frac{\partial X}{\partial x}(0)Y(0)\right) + \dots \tag{24}$$

where the remaining terms are all of higher order in t and thus vanish for the second derivative evaluated at zero. Differentiating and evaluating at $t = 0$ yields

$$\frac{d^2}{dt^2}\bigg|_{t=0} \left(\phi_{-t}\theta_{-t}\phi_t\theta_t\right)(0) = 2\left(\frac{\partial Y}{\partial x}(0)X(0) - \frac{\partial X}{\partial x}(0)Y(0)\right) \tag{25}$$

```
In[21]:= ClearAll["Global`*"]
         (*X is X(0) and A is the spatial derivative of X, (dX/dx)(0).
           Higher order terms excluded as they will vanish upon evaluation.*)
         theta[t_, x_] := x + t * (X + A * x) + ((t^2) / 2) * A * X;
         (*Y is Y(0) and B is the spatial derivative of Y, (dY/dx)(0).
           Higher order terms excluded as they will vanish upon evaluation.*)
         phi[s_, y_] := y + s * (Y + B * y) + ((s^2) / 2) * B * Y;
         (*Defining the commutator of the flows*)
         comm[s_, t_] := phi[-s, theta[-t, phi[s, theta[t, 0]]]];

         (*Takes the derivative of the commutator with respect to t
           then evaluates at t=0*)
         Expand[comm[t, t]]

         D[D[Expand[comm[t, t]], t], t] /. t → 0
```

Out[25]= $B\,t^2\,X - \frac{1}{2}A^2\,t^3\,X - \frac{1}{2}A\,B\,t^3\,X - B^2\,t^3\,X + \frac{1}{2}A\,B^2\,t^4\,X + \frac{1}{2}A^2\,B^2\,t^5\,X - A\,t^2\,Y + \frac{1}{2}A\,B\,t^3\,Y - \frac{1}{2}B^2\,t^3\,Y + \frac{1}{2}A\,B^2\,t^4\,Y$

Out[26]= $2\,B\,X - 2\,A\,Y$

Figure 6: Mathematica code for calculating the second derivative of the commutator.

Up to the factor of two, the right hand side in Equation 25 in coordinates is

$$\frac{\partial Y}{\partial x}(0)X(0) - \frac{\partial X}{\partial x}(0)Y(0) = \left(\frac{\partial Y^j}{\partial x^i}(0)X^i(0) - \frac{\partial X^j}{\partial x^i}(0)Y^i(0)\right)e_j \tag{26}$$

which is the form of Equation 15 above. This concludes the proof.

It is perhaps worth noting that you may raise objections to this in so far as we're taking second order derivatives. Certainly the first derivative of a curve $\gamma : I \longrightarrow M$ lies in the tangent space $T_{\gamma(0)}M$, but not the second. The reason why we may consider the second derivative to be an element of the tangent space is extremely subtle. Let's say you have a curve $\gamma : I \longrightarrow M$ such that $\gamma'(0) = 0$. In general tangent vectors of curves act on the algebra of smooth functions $C^\infty(M)$ via $\gamma'(0)f = (f \circ \gamma)'(0)$ where $f \circ \gamma : I \longrightarrow \mathbb{R}$ is a smooth function. Because the first derivative vanishes, we may actually define an action of $\gamma''(0)$ on $C^\infty(M)$ via $\gamma''(0)f = (f \circ \gamma)''(0)$. In general this does not satisfy the criterion of being a derivation, but because $\gamma'(0) = 0$, it does. It is worth writing out the technical details to ensure that this is so, but once this is verified, we may indeed identify $\gamma''(0) \in T_pM$. In our case, one can see via Equation 24 that this condition is satisfied, as the commutator possesses no first order terms.

References

[1] Timothy Goldberg. What is a connection and what is it good for? 2008. URL `http://pi.math.cornell.edu/~goldberg/Notes/AboutConnections.pdf`.

[2] John Lee. Smooth Manifolds Second Edition. pages 227–233, 2000.