\[(f^*\alpha)_p(x) = \alpha_{f(p)}(f^*x), \quad \forall \ p \in M\]

where the subscript indicates the point at which the field is evaluated. Thus 1-forms on \(N\) get pulled back into 1-forms on \(M\).

A simpler example of the pull-back is for scalar fields. Let \(\phi \in \mathcal{F}(N)\), \(\phi : N \to \mathbb{R}\) be a scalar field. Then we define the pull-back \(\psi = f^*\phi \in \mathcal{F}(M)\) by

\[\psi(p) = (f^*\phi)(p) = \phi(f(p)) = (\phi \circ f)(p),\]

that is, \(f^*\phi = \phi \circ f\).

Return to the pull-back of vectors, and put it into coordinate language. Let \(x^i, y^i\) be coordinates on \(M\) and \(N\) as above, and let \(\beta = f^*\alpha\), so that

\[\alpha = \sum_{i=1}^{n} \alpha^i \ dy^i|_p\]

\[\beta = \sum_{i=1}^{m} \beta^i \ dx^i|_p\]

Then

\[\beta^i = \sum_{j=1}^{n} \frac{\partial y^i}{\partial x^j} \alpha^j.\]

Notice that \(f^{-1}\) need not be defined in order to define \(f^*\) and \(f^*\). In particular, \(M\) and \(N\) need not have the same dimensionality. But if \(f^{-1}\) does exist (then \(\dim M = \dim N\)), then we can "push forward" vectors from \(M\) to \(N\) (by \(f^{-1}\ast\)) and

(continued page after next)
Behavior of $f^*$, $f_*$ under compositions. Let $f : M \to N$, $g : N \to P$.

Then $g \circ f : M \to P$,

and $(g \circ f)_* = g_* \circ f_*$. 

This is fairly obvious, you just map the small displacement vector (in $M$) under a succession of 2 linear maps, first by $f_*$, then $g_*$, to get $(g \circ f)_*$.

As for pull-backs, the rule is

$$(g \circ f)^* = f^* \circ g^*$$

(in reverse order).
"pull-back" vectors from \( N \) to \( M \) (by \( f^* \)).

Now consider the mapping of one manifold of a certain dimensionality into one of higher dimensionality, \( f: M \to N \), \( \dim M \leq \dim N \). Then \( f \) is called an immersion if \( f_* \) is of maximal rank.

\[
\text{rank } f_* : T_p M \to T_{f(p)} N = \dim M.
\]

This means that each little piece of \( M \), which looks like \( \mathbb{R}^m \), gets mapped into a subset of \( N \) that also looks like \( \mathbb{R}^m \). This means that \( f_* \) is injective (the image of \( M \) under \( f \) is locally \( m \)-dimensional).

However, an immersion does not preclude self-intersections:

\[\begin{array}{c}
\text{To exclude self-intersections, we can demand that } f \text{ itself be an injection. This means that } f \text{ is called an embedding (because } m \text{ of } f \text{ "looks like" } M). \\
\end{array}\]

Now we consider ordinary differential equations (ODE's) and flows. Begin with an intuitive picture of a vector field, as a small displacement (each understood to be taking place in some elapsed parameter \( t \)), attached to each point of \( M \):
picture of $X \in \mathcal{X}(M)$

Idea: Each point gets up and moves a small amount in time $\Delta t$.

If you just follow these arrows, starting with some initial point $x_0$, you trace out a curve called the integral curve of $X$. By following an integral curve for time $t$, starting at $x_0$, you get a final point described by a function

$$\Phi: M \times \mathbb{R} \rightarrow M,$$

$$x = \Phi(x_0, t),$$

where $\Phi$ is called the advance map.

To make this more precise, express the vector field $X$ in some chart:

$$X = \sum_i X^i(x) \frac{\partial}{\partial x^i}$$

This is an operator which when acting on scalars $f$ gives a number interpreted as $df/dt$. In particular, letting it act on the coordinates themselves gives a set of ODEs:

$$\frac{dx^i}{dt} = X^i(x).$$

Thus, a vector field on a manifold is a generalization of a system of ODE's on $\mathbb{R}^n$. Standard theorems on ODE's say that the system above has a unique solution $x(t)$ satisfying $x(0) = x_0$, $f$ for $t$ in some interval.
[0,1], if the functions $x^i(x)$ are smooth. This is the (important) uniqueness theorem for ODE's (really, existence and uniqueness).

However, even if the vector field $x^i(x)$ is smooth, the solution may not exist for all $t$. (For example, it may run off to infinity in finite $t$. (For an example of this, consider $x = x^1, x_0 = 1$ ($x \in \mathbb{R}$), for which $x \to \infty$ at $t \to 1$.) For simplicity, we will assume that this does not happen, i.e., that solutions $x^i(t)$ exist for all time, for any $x_0$. Then we can speak of the "general solution functions" $\Phi^i(t, x_0)$ that give $x^i(t)$, assuming $x^i(0) = x_0^i$. These solution functions satisfy

1) $\Phi^i(0, x_0) = x_0^i$

2) $\frac{d\Phi^i}{dt}(t, x_0) = x^i(\Phi(t, x_0))$.

These are just the initial conditions and ODE's expressed in terms of $\Phi^i$. [We would normally write them,

1) $x^i(0) = x_0^i$

2) $\frac{dx^i}{dt} = x^i(x(t))$. ]

All of the above is in one chart. By mapping a solution $x^i(t)$ in the given chart back onto $M$, we get a segment of an integral curve. [Just before we run off one chart we can switch to another, thereby continuing the integral curve. The result is that we define a map $\Phi : \mathbb{R} \times M \to M$ or maps $\Phi_t : M \to M$. (a different notation), such that]
\( x = \Phi(t, x_0) = \Phi_t(x_0) \)

is the point on the integral curve starting at \( x_0 \) at \( t=0 \), reached after time \( t \).

The advance map satisfies an important property,

\[
\Phi_s \Phi_t = \Phi_{s+t}
\]

or \( \Phi(s, \Phi(t, x_0)) = \Phi(st, x_0) \).

(the composition property). This is intuitive: if you start at \( x_0 \),
follow the integral curve for elapsed time \( t \), reaching \( x_1 \), then
reatt \( x_1 \) as initial conditions and follow the integral curve for elapsed
time \( s \), you must get the same thing as starting at \( x_0 \) and following
the integral curve for time \( s+t \).

We will prove this working in a single chart, ignoring the
complications that result when we must switch charts. We want to
show that

\[
\Phi^i(s, \Phi^i(t, x_0)) = \Phi^i(s+t, x_0).
\]

Let \( x_1^i = \Phi^i(t, x_0) \), \( \xi^i(s) = \Phi^i(s, x_1) \), \( \eta^i(s) = \Phi^i(s+t, x_0) \).

We need to show that \( \xi^i(s) = \eta^i(s) \). First, at \( s=0 \), we have

\[
\xi^i(0) = \Phi^i(0, x_1) = x_1^i
\]

\[
\eta^i(0) = \Phi^i(t, x_0) = x_1^i.
\]

Next, we have

\[
\frac{d \xi^i}{ds} = \frac{\partial \Phi^i}{\partial s}(s, x_1) = \Phi^i(\Phi(s, x_1)) = \Phi^i(\xi(s)),
\]
and
\[
\frac{d\eta^i}{ds} = \frac{\partial \Phi^i(s+t, x_0)}{\partial s} = \frac{\partial \Phi^i}{\partial s}(s+t, x_0)
\]

\[
= X^i(\Phi(s+t, x_0)) = X^i(\eta(s)).
\]

Thus, both \(\Phi^i(s)\) and \(\eta^i(s)\) satisfy the same ODEs and the same initial conditions, so by the uniqueness theorem they must be equal. QED.

By the composition property, \(\Phi_t \circ \Phi_t = \text{id}_M\), so \(\Phi_t\) is a diffeomorphism \(M \rightarrow M\). In fact, the set

\[
\{\Phi_t \mid t \in \mathbb{R}\}
\]

constitutes a one-parameter group of diffeomorphisms of \(M\) onto itself. It is an action of the group \(\mathbb{R}\) (meaning \(+\)) on \(M\). This group is sometimes called the flow.

Now about the exponential notation for the flow. This is a way of connecting \(\Phi_t\) with the vector field \(X\). The notation used in many books is

\[
\Phi_t = e^{tX}.
\]

Taken as written, this has no meaning (we must assign a meaning to it). First note that a vector field \(X\) is a mapping \(\mathcal{F}(M) \rightarrow \mathcal{F}(M)\), namely, \(f \mapsto \frac{\partial f}{\partial x^i} X^i\). Therefore \(X^2 = X \cdot X = X \circ X\) has a meaning as a map \(\mathcal{F}(M) \rightarrow \mathcal{F}(M)\). The higher powers of a vector field are not vector fields (they are not 1st order partial differential operators, they are higher order diff. ops.), but they are perfectly good maps \(\mathcal{F}(M) \rightarrow \mathcal{F}(M)\). [Notice that a vector at a point is a map \(\mathcal{F}(M) \rightarrow \mathbb{R}\), so a power of it has no meaning.]
So, an exponential series like

\[ e^{tX} = 1 + tX + \frac{t^2}{2!} X^2 + \ldots \]

has meaning as a map: \( \Phi(t) : \mathcal{F}(M) \rightarrow \mathcal{F}(M) \), at least if we ignore convergence questions (which we will). The \( 1 \) above means \( \text{Id}_M \). On the other hand, \( \Phi(t) \) is a map: \( M \rightarrow M \), not: \( \mathcal{F}(M) \rightarrow \mathcal{F}(M) \). This is why \( \Phi(t) = e^{tX} \) has no meaning as it stands. But let us apply the exponential series to a function \( f \) and see what we get. Let us evaluate the function at \( x_0 \), which is an initial condition. We start with the \( t \)-expansion of \( e^{tX} f \):

\[ e^{tX} f = f + tXf + \frac{t^2}{2} X^2 f + \ldots \]

\[ = f + t \sum_i x_i \frac{\partial}{\partial x_i} f + \frac{t^2}{2} \sum_i x_i \frac{\partial}{\partial x_i} \sum_j x_j \frac{\partial}{\partial x_j} f + \ldots \]

\[ = f + t \ " \frac{df}{dt}" + \frac{t^2}{2} \ " \frac{d^2f}{dt^2}" + \ldots \]

where we put the \( t \)-derivatives in quotes because what is actually meant is

\[ \frac{df}{dt} = \frac{d}{dt} (f \circ c) \]

where \( c \) is the integral curve passing through \( x \) at \( t=0 \). Thus, if the series converges, it does so to \( " f(t)" \), which means \( f(x(t)) \), where \( x(t) \) is the integral curve, \( x(t) = \Phi(t)(x_0) \). So, \( \Phi(t) \) means the same as \( c(t) \)

\[ (e^{tX} f)(x_0) = f(\Phi(t)(x_0)) = (\Phi^*_t f)(x_0). \]

This is true for all \( x_0 \), so we have

\[ e^{tX} f = \Phi^*_t f. \]
And this is true for all $t$, so we have
\[ \Phi_t^* = e^{tX}. \quad (A) \]

The usual formula in books is meaningful if we put a $*$ on $\Phi_t$ (turning it into a pull-back), and interpret both sides as maps $\mathcal{F}(M) \to \mathcal{F}(M)$.

On the other hand, we can regard $e^{tX}$ as a formal notation for $\Phi_t$, without trying to interpret it as a power series. This notation has the virtue of making some of the properties of the advance map obvious:
\[ \Phi_t \Phi_t^* = e^{tX} e^{tX} = e^{2tX} = \Phi_{2t}, \]
and $\Phi_0 = e^{0X} = 1 = \text{id}_M$.

Restoring the star $*$, we can differentiate $\Phi_t^* = e^{tX}$ formally, to get
\[ \frac{d}{dt} \Phi_t^* = xe^{tX} = e^{tX}x, \]
which implies
\[ \frac{d}{dt} \Phi_t^* = x\Phi_t^* = \Phi_t^*x, \quad (B) \]
an equation that is perfectly meaningful as operators $\mathcal{F}(M) \to \mathcal{F}(M)$. We have not proved it (because of questions of convergence of series), but in fact this result is true, and can be proved by other means.

Altogether, we have derived relationships between a vector field $X$ and the 1-parameter group of diffeomorphisms $\{\Phi_t\}$ that it generates, in an intrinsic notation (not tied to a coordinate system). These are Eqs. (A) and (B) above.
Continue today with the Lie derivative, which is like the convective derivative of ordinary tensor analysis, but generalized to arbitrary manifolds.

Context: Given a manifold $M$, a vector field $X \in \mathfrak{X}(M)$, with advance map $\Phi_t : M \to M$. Illustrate Lie derivative first with scalar fields, where $L_X : \mathcal{F}(M) \to \mathcal{F}(M)$ ($L_X$ is the Lie derivative along vector field $X$)

Let $x_0 \in M$ and $x_1 = \Phi_t x_0$. We think of $t$ as small (we will be interested in the limit $t \to 0$). For $f \in \mathcal{F}(M)$, define

$$(L_X f)(x_0) = \lim_{t \to 0} \frac{1}{t} \left[ f(x_1) - f(x_0) \right]$$

Integral curve of $X$.

$x_1 = \Phi_t(x_0)$

It's pretty obvious from this formula that $L_Xf = Xf$, since the vector $X|_{x_0}$ is the small displacement $x_0 \to x_1$ in small time $t$.

Thus, the Lie derivative of a scalar is the obvious generalization of the convective derivative to an arbitrary manifold,

$$L_X f = Xf = \sum_i X^i \frac{\partial f}{\partial x^i} \quad \text{(think $\nabla \cdot Xf$)}.$$

But transform eqn. above:

$$(L_X f)(x_0) = \lim_{t \to 0} \frac{1}{t} \left[ f(\Phi_t x_0) - f(x_0) \right]$$

$$= \lim_{t \to 0} \frac{1}{t} \left[ (\Phi_t^\ast f)(x_0) - f(x_0) \right]$$
\[
\begin{align*}
\frac{d}{dt} \Phi_t^x \bigg|_{t=0} &= \frac{1}{t} \left( \left( \Phi_t^x \right) |_{t=0} - 1 \right) f(x_0) \\
&= \left( \left( \frac{d}{dt} \Phi_t^x \right) |_{t=0} \right) f(x_0).
\end{align*}
\]

But recall,
\[
\Phi_t^x = e^{tx}
\]
when acting on scalars, so we find again,
\[
\frac{d}{dt} \Phi_t^x \bigg|_{t=0} = x, \quad L_x f = xf.
\]

Now generalize to other differential geometric objects, like vector fields. Now we want to define \( L_x \) as an operator: \( \mathfrak{X}(M) \rightarrow \mathfrak{X}(M) \).

Let \( Y \in \mathfrak{X}(M) \) (now we have 2 vector fields, \( X \) and \( Y \)), and we wish to define \( L_x Y \). The idea is the same as above, we wish to compare \( Y \) at \( x_1 \), with \( Y \) at \( x_0 \), to see how much \( Y \) has changed along the integral curves of \( X \). But we cannot just subtract

\[
Y(x_1) - Y(x_0), \quad \text{these vectors belong to two different tangent spaces} \quad T_{x_1}M \quad \text{and} \quad T_{x_0}M \quad \text{without any natural identification.}
\]

However, we can "pull-back" \( Y(x_1) \) to point \( x_0 \) using the flow (mapping both base and tip of arrow by \( \Phi_t^x \)). Note that the pull-back of a vector field is defined in this case because \( \Phi_t^x \) is invertible: the pull-back is the inverse of the tangent map \( \Phi_t^x \). So, we define
\[ \mathcal{L}_x Y = \left( \frac{d}{dt} \bigg|_{t=0} \Phi^{-1}_{t,x} \right) Y, \quad \text{or} \]

\[ (\mathcal{L}_x Y)(x_0) = \lim_{t \to 0^+} \left[ (\Phi^{-1}_{t,x} Y)(x_0) - Y(x_0) \right] \]

In components,

\[ (\Phi^{-1}_{t,x} Y)^i(x_0) = \frac{\partial x_i^j}{\partial x_j^i} Y^j(x_1). \]

To get \( x_1 \) as a fun of \( x_0, t \), we solve the ODE's,

\[ \frac{dx_i^t}{dt} = X_i^t(x) \]

in power series in \( t \),

\[ x_i^t = x_i^0 + t X_i^t(x_0) + \ldots \]

or its inverse,

\[ x_0 = x_i^0 - t X_i^t(x_0) + \ldots \]

So,

\[ \frac{\partial x_i^0}{\partial x_j^i} = \delta_j^i - t X_{i,j}^i + \ldots \]

and,

\[ \rightarrow = \left[ \delta_j^i - t X_{i,j}^i \right] Y^j(x_0 + t X + \ldots) \]

\[ = Y^i(x_0) + t \left( X^j Y_{j,i}^i - Y^j X_{i,j}^i \right). \]

So,

\[ (\mathcal{L}_x Y)^i = X^j Y_{j,i}^i - Y^j X_{i,j}^i. \]

This is the Lie derivative of a vector field.

Similarly, one can define the Lie derivative of a covector field.
by

\[ L_x \alpha = \left( \frac{d}{dt} \bigg|_{t=0} \Phi_t^* \right) \alpha. \]

If you work it out, you find (in components),

\[ (L_x \alpha)_i = X^j \alpha_{i,j} + \alpha^i X_j, \]

To define \( L_x \) on arbitrary tensors, we develop some general rules. First, \( L_x \) acts on a tensor product of tensors by the Leibnitz rule. An example will illustrate. Consider the tensor product of a covector with a vector (this is a type \((1,1)\) tensor): Define

\[ L_x (\alpha \otimes Y) = \frac{d}{dt} \bigg|_{t=0} \left[ (\Phi_t^* \alpha) \otimes (\Phi^{-1}_t Y) \right], \]

which is the obvious definition. But this is...

\[ \Rightarrow (L_x \alpha) \otimes Y + \alpha \otimes (L_x Y), \]

(Leibnitz rule).

The same thing works for contractions. For example, the tensor \( \alpha \otimes Y \) has components,

\[ (\alpha \otimes Y)_i^j = \alpha_i Y^j. \]

If we contract (set \( i=j \) and sum), we get

\[ \alpha_i Y^i = \alpha(Y) = \text{a scalar}. \]

Then we have

\[ L_x [\alpha(Y)] = (L_x \alpha)(Y) + \alpha (L_x Y) \]

\[ = \alpha(Y). \]

Can use this to calculate \( L_x \alpha \) in components, supposing that we know what \( L_x \) does to scalars.
and vector fields.

Notice that a scalar multiplied by a tensor is a special case of a tensor product:

\[ f \otimes T = fT \quad \text{for any } T, \quad f \in \mathcal{F}(\mathcal{M}). \]

Therefore,

\[ \mathcal{L}_X (fT) = (\mathcal{L}_X f) T + f(\mathcal{L}_X T) = (Xf) T + f(\mathcal{L}_X T). \]

Since an arbitrary tensor can be written as linear combinations of scalars times tensor products of vector fields and covector fields, the Leibnitz rule suffices to compute the Lie derivative of any tensor.

Some more rules about \( \mathcal{L}_X \). If \( f \in \mathcal{F}(\mathcal{M}) \), then

\[ \mathcal{L}_X f \]

This is obvious since \( f X \) has some integral curves as \( X \), except the \( t \)-parametrization is scaled by \( f \). Hence \( \frac{\partial}{\partial t} \bigg|_{t=0} \) is scaled by \( f \).

The Lie derivative \( \mathcal{L}_X Y \) is a special case, with a special interpretation. Consider the flows associated with \( X, Y \), call them \( \Phi_s, \Psi_t \). These in general do not commute,

\[ \Phi_s \Psi_t \neq \Psi_t \Phi_s. \]

Small vector field, when \( s, t \) small.
When $s,t$ are small, the difference in the endpoints must be a vector. However, since we cannot subtract points, to measure the difference between $\Psi_0 \Phi_s x_0$ and $\Phi_s \Psi_0 x_0$ we evaluate some scalar $f : M \to \mathbb{R}$ at the 2 points and subtract:

$$f(\Psi_0 \Phi_s x_0) - f(\Phi_s \Psi_0 x_0)$$

$$= (\Psi_0 \Phi_s)^* f(x_0) - (\Phi_s \Psi_0)^* f(x_0)$$

$$= (\Phi_s^* \Psi_0^* f)(x_0) - (\Psi_0^* \Phi_s^* f)(x_0)$$

$$= \left( (\Phi_s^* \Psi_0^* - \Phi_s^* \Phi_0^* \Psi_0^*) f \right)(x_0).$$

$$e^{sx + ty} e^{-sx} e^{tx} e^{-ty}$$

$$= (1 + sx + \frac{s^2}{2} x^2 + \ldots)(1 + ty + \frac{t^2}{2} y^2 + \ldots) - (x \otimes y)$$

$$= 1 + (sx + ty) + \left(\frac{s^2}{2} x^2 + stxy + \frac{t^2}{2} y^2 + \ldots\right) - (x \otimes y)$$

$$= st \left( xy - yx \right) + \ldots$$

Thus the small vector is $[x,y]$ times $st$.

Thus the leading term is the commutator of $X$ and $Y$ (regarded as maps: $F(M) \to F(M)$). Thus we have

$$\frac{\partial^2}{\partial s \partial t} \bigg|_{t=s=0} \left[ f(\Psi_s \Phi_t x_0) - f(\Phi_t \Psi_s x_0) \right] = ([x,y] f)(x_0),$$

for all $f \in F(M)$. 