A Self-Contained Introduction to Lie Derivatives

Ebrahim Ebrahim

May 19, 2010

Contents

0.1 Introduction

The goal of this set of notes is to present, from the very beginning, my understanding of Lie derivatives. I delve into greater detail when I do topics that I have more trouble with, and I lightly pass over the things I understand clearly. That is, these are more like personal notes than they are like a textbook. Some gaps would need to be filled in if this were to be used for teaching (and someday I may fill in these gaps). Although I skip proofs for certain things in the interest of time, I make sure to note what I've skipped for the reader. If you want to jump to the end of these notes because you're already familiar with the basics of differential geometry, then make sure you check the notation part of the appendix. Also anyone may easily skip the first chapter. It didn't flow into the rest of the paper as well as I'd hoped. I'd probably have to double the length of that chapter to connect it to everything else, but it isn't worth it.

I admit that I'm very wordy in my explanations here. If I just wanted to present a bunch of definitions this wouldn't be a very useful document. Any textbook serves that purpose. My goal here is to convey my own thoughts about the topics I present. Aside from things like topology theorems or the tensors section, this is very much my personal take on the subject. So please excuse the wordiness, and if you just want an explanation that gets to point this is not the thing to read.

I use the following books:

- Warner's "Foundations of Differentiable Manifolds and Lie Groups"
- Bishop and Goldberg's "Tensor Analysis on Manifolds"
- Bernard Schutz's "Geometrical Methods of Mathematical Physics"
- Frankel's "The Geometry of Physics"

On a scale from physics to math, I would rate these authors like this: Schutz, Frankel, Bishop and Goldberg, Warner. Bishop and Goldberg was the most practical book. Warner is a difficult read, but it is the most mathematically honest (and my personal favorite). Schutz does a great job developing intuitive concepts, but Schutz alone is absolutely unbearable. Having gone through a good amount of these books, here are my recommendations:

For the physicist: Schutz with Bishop and Goldberg

For the mathematician: Warner with Bishop and Goldberg

Frankel is more of a reference book. It has good explanations but it doesn't flow in a very readable order.

Chapter 1

Preliminaries

I decided to mostly follow Bishop's treatment for this section. It was the most rigorous set of preliminaries that wasn't excessively detailed. It is a lighter version of the one by Warner. Serge Lang has a very nice introduction that deals with category theory and topological vector spaces, but it's not necessary for these notes. Frankel likes to do things a little out of order; he always motivates before defining. It's good for a first time reading, but not for building a set of notes.

I assume a basic knowledge of topology, so I go through the definitions as a quick overview, this is not intended to be read by a first-timer. This section is kind of disconnected from the succeeding sections since I chose to avoid spending all my time on manifolds and their topology. This section may be skipped without any problems for now.

1.1 General Definitions

Definition 1.1.1. *A topological space is a pair, (X, τ), consisting of an underlying set and a topology. The underlying set is commonly referred to as the topological space, and the topology must be a set of subsets of the topological space which is closed under arbitrary unions and finite intersections.*

Topological spaces are sets endowed with a very bare structure that just barely gives them the privilege of being called spaces. The topology contains the set of subsets of the space that we consider to be open. So a topological space is a set for which we have given a meaning to the word "nearby." A topological isomorphism is called a *homeomorphism.*

Definition 1.1.2. *A homeomorphism between a topological space, (X, τ_X), and a topological space, (Y, τ_Y), is a bicontinuous bijection (a continuous 1-1 onto map with a continuous inverse) from X to Y.*

How is this a proper definition of homeomorphism? Well a topological isomorphism should take one space onto another and preserve openness. That is, an element of the topology of one space should have an image under the homeomorphism which is an element of the topology of the other space, and vice versa. So if $f : X \overset{bij}{\to} Y$ is a homeomorphism, then $S \in \tau_X$ iff $f[S] \in \tau_Y$. This is in fact the case when we define continuity. A function from one topological space to another is *continuous* iff the inverse image of any open set in the range is open.

If we take a subset A of a topological space, (X, τ_X), the topological *subspace* induced by it has the topology $\{G \cap A | G \in \tau_X\}$.

A more direct but less general way to give a set this structure is through a metric, a distance function. Now this use of the word "metric" is *not* the same as the metric of a manifold in Riemannian geometry. This is the metric of a *metric space*, do not confuse the two.

Definition 1.1.3. *A metric space is a pair, (X, d), consisting of an underlying set and a distance function (or metric). The distance function, $d : X \times X \to \mathbb{R}$, must be positive, nondegenerate, symmetric, and it must satisfy the triangle inequality. The underlying set is commonly referred to as the metric space.*

Positivity means it always gives positive distance, nondegeneracy means that $d(x, y) = 0 \Leftrightarrow x = y$, symmetry means that $d(x, y) = d(y, x)$, and the triangle inequality means that $d(x, y) + d(y, z) \geq d(x, z)$. All metric spaces can be made into topological spaces in the obvious way (use the set of open balls as a base), but not all topological spaces are metrizable.

Definition 1.1.4. *A topological space X is Hausdorff if any pair of distinct points has a corresponding pair of disjoint neighborhoods. That is, $(\forall x, y \mid x, y \in X \Rightarrow (\exists G, H \mid G, H \in \tau_X \bullet G$ neighborhood of $x \bullet H$ neighborhood of $y \bullet G \cap H = \{\}))$.*

Hausdorff spaces are usually pretty normal, they are all we care about in physics. Metric topologies are always Hausdorff. Singletons, $\{x\}$, are always closed in Hausdorff spaces.

$\mathbb{R}^n$ is the set of n-tuples of real numbers. A real n-tuple is a finite length-n sequence of real numbers. Typically whenever we mention $\mathbb{R}^n$ we immediately assume that it is equipped with the "standard topology." This is the topology induced by the standard metric, $d(x, y) = \sqrt{\sum_{i=0}^{n} (x_i - y_i)^2}$.

1.2 Connected Sets

We define connectedness in terms of non-connectedness.

Definition 1.2.1. *A topological space X is not connected iff there are nonempty sets G, H such that $G \cap H = \{\}$ and $G \cup H = X$.*

Theorem 1.2.1 (Chaining Theorem)**.** *If $\{A_a | a \in J\}$ is a family of connected subsets of X and $\bigcap_{a \in J} A_a \neq \{\}$ then $\bigcup_{a \in J} A_a$ is connected.*

Proof. Assume the hypotheses. Suppose $\bigcup_{a \in J} A_a$ is not connected. Get G, H such that $G \cap H = \{\}$ and $G \cup H = \bigcup_{a \in J} A_a$. We have $\bigcap_{a \in J} A_a \neq \{\}$, so get $x \in \bigcap_{a \in J} A_a$. x is either in G or it's in H, say it's in G. Since H is not null, get $a \in J$ such that $A_a \cap H \neq \{\}$. Then $G \cap A_a$ and $H \cap A_a$ are disjoint nonempty open sets who do not meet and whose union is A_a. This contradicts that A_a is connected. □

This was a sample of the kinds of theorems that one would deal with when handling connectedness (a particularly easy one at that). The most important thing that should be clear is that connectedness is a topological property, it is preserved under homeomorphisms. I'll just provide a brief synopsis of some other theorems, they were worth studying but they take too long to write up: The closure of a connected set is connected. For continuous functions, connected subsets of the domain have connected images (a generalization of intermediate value theorem). An *arcwise connected* topological space has the property that any two points in it can be connected by a continuous curve in the space, this is more strict than the condition for connectedness. Bishop and Goldberg additionally show that a topological space can be reduced to maximally connected components. These are all interesting but not crucial results for the purposes of these notes.

1.3 Compactness

For $A \subset X$, a *covering* of A is a family of subsets of X whose union contains A. When the subsets are all open it's called an *open covering.* A *subcovering* of a covering $\{C_a \mid a \in I\}$ is another covering where the index set is just restricted, $\{C_a \mid a \in J\}$, $J \subset K$. When the index set is finite it's called a *finite* covering. Compactness is also a topological property, it is preserved under homeomorphisms. The Heine-Borel theorem for $\mathbb{R}$ generalizes to $\mathbb{R}^n$ and tells us that closed bounded subsets of $\mathbb{R}^n$ are compact. I will not prove all of the following theorems, the proofs can be found in Bishop and Goldberg on page 16.

Theorem 1.3.1. *Compact subsets of Hausdorff spaces are closed.*

Theorem 1.3.2. *Closed subsets of compact subspaces are compact.*

Proof. Consider a closed subset of a compact space. Take the complement of the closed subset, this can be added to any open covering of the closed subset to get an open covering of the whole compact space. Now there exists a finite subcovering of the whole space, and the complement of the closed subset can now be removed to leave a finite subcovering of the closed subset. □

Theorem 1.3.3. *Continuous functions have maxima and minima on compact domains.*

Theorem 1.3.4. *A continuous bijection from a compact subspace to a Hausdorff space is a homeomorphism.*

A topological space is *locally compact* if every point of it has a compact neighborhood (compact spaces are then locally compact). A topological space is *separable* if it has a countable basis. Another word for this is *second countable.* A family of subsets of a topological space is *locally finite* if every point in the space has a neighborhood that touches a finite number of subsets. A covering A_a of a topological space X is a *refinement* of the covering B_b if for every index a there is a set B_b such that $A_a \subset B_b$. What was the point of all that gobbledygook? Well now we can finally define *paracompactness*; a topological space is paracompact if every open cover has an open locally finite refinement. Some authors require that a space be Hausdorff to be paracompact (Bishop and Goldberg 17-18), and others do not (Warner 8-10). The ones that do not just have slightly harder proofs, which can be found on the indicated pages. The bottom line is the following theorem.

Theorem 1.3.5. *A locally compact second-countable Hausdorff space is paracompact.*

Why in the world would we need such a thing? Well it just so happens that manifolds have these very properties.

Chapter 2

Manifolds

Theoretically oriented books on differential geometry are rich with theorems about manifolds. Since these notes are geared towards building a foundation to do physics, I will be more interested in definitions and explanations than pure theorems. I do, however, want to be sure that the definitions presented are completely precise and coherent. For this reason I chose to use Warner's treatment. I will supplement this with my own examples and explanations. If I ever get around to adding them, then for my own reference here is a short list of the things I've left out: inverse function theorem, submanifolds, immersions and imbeddings, manifolds with boundaries, conditions for smoothness of curves in weird cases, and some application of the topology theorems from the previous section.

2.1 Definitions

Definition 2.1.1. *A function on $\mathbb{R}^d$ with an open domain in $\mathbb{R}^n$ is said to be differentiable of class C^k if all its partial derivatives of order less than or equal to k are continuous for all it's component functions.*

This terminology is used to set the level of differentiability available. The term *smooth* is used for C^∞, differentiability of all orders. In particular, a C^0 function is continuous.

Definition 2.1.2. *A locally euclidian space M of dimension d is a Hausdorff topological space in which each point has a neighborhood homeomorphic to an open subset of $\mathbb{R}^d$.*

The homeomorphisms of a locally euclidean space are called *coordinate systems*, *coordinate maps*, or *charts* (e.g. $\phi : U \to \mathbb{R}^d$, $U \subset M$). Their component functions, each a map from the open set in M to $\mathbb{R}$, are called *coordinate functions.*

Definition 2.1.3. *A differentiable structure of class C^k, $\mathscr{F}$, on a locally euclidean space M of dimension d is a set of coordinate maps (charts) such that:*

- *The domains of the coordinate systems cover M.*
- *All compositions of coordinate maps $\phi_a \circ \phi_b^{\leftarrow}$ are C^k differentiable.*

- *$\mathscr{F}$ is maximal with respect to the first two properties. That is, any coordinate system with the above properties is in $\mathscr{F}$.*

A set of coordinate maps satisfying only the first two properties is usually just called an *atlas*. A differentiable structure is then just a maximal atlas. The reason we require the differentiable structure to be maximal is because we could otherwise have two *unequal* manifolds that differ only in the particular choice of atlas.

Definition 2.1.4. *A d-dimensional differentiable manifold of class C^k is a pair $(M, \mathscr{F})$ where M is a d-dimensional, second countable, locally euclidean space and $\mathscr{F}$ is a differentiable structure of class C^k on M.*

And we finally have the definition of a manifold. Most authors avoid getting into differentiability classes throughout their books. They typically comment at the beginning that one may assume either smoothness or as much differentiability as is needed for a particular discussion. There are also analytic manifolds and complex manifolds, but we will not get into them.

Now that we've defined manifolds, we can immediately start thinking about doing calculus on them. Differentiability of functions is quite simple. Each coordinate system is sort of like a pair of glasses through which one may view the manifold and the things living on it. If I were to put, say $f : U \to \mathbb{R}$ ($U \subset M$), on the manifold, then to view the function through the eyes of a coordinate map ϕ I just have to look at $f \circ \phi^{\leftarrow}$. This is a map from $\mathbb{R}^d$ to $\mathbb{R}$, where d is the dimension of the manifold. This is an object on which we know how to do regular calculus. We use this object to define C^k differentiability for functions on the manifold: A function f from a manifold to the real numbers is C^k *differentiable* iff $f \circ \phi^{\leftarrow}$ is C^k differentiable for all coordinate maps ϕ. If $\psi : M \to N$, where M, N are manifolds, then ψ is C^k differentiable iff $\tau \circ \psi \circ \phi^{\leftarrow}$ is C^k differentiable for all coordinate maps ϕ on M and τ on N.

Notice that the definition of a manifold makes it paracompact, straight from theorem 1.3.5. This is an important result for mathematicians, but we will not be getting into it too deeply. There are other important theorems we won't be using, like the existence of partitions of unity, just because our goal is to apply this to general relativity.

2.2 Examples

If V is a finite dimensional vector space, we can make a nice manifold out of it. Given a basis, a dual basis *is* the coordinate functions of a coordinate map on all of V. It's C^∞ too! This example is particularly interesting because V also has the group properties of a vector space, which makes it a Lie group. (Read this statement again after reading section 4.2 if you don't get it).

The sphere can be made into a manifold if we chart it using, for example, the typical θ and ϕ coordinates. Notice, however, that the domain of a coordinate map must be an open set (it has to be a homeomorphism after all). This makes it impossible to put a global coordinate system on a sphere, more than one coordinate map is necessary to cover it. Another interesting set of charts is stereographic projection, and yet another one is projecting the six hemispheres to six open disks.

I'd like to present one very concrete example of a manifold before moving on. The torus as a topological space is homeomorphic to a half-open square:

$$Torus = \{(x, y) \mid 0 \leq x < 1 \bullet 0 \leq y < 1\}$$

Now we just have to put coordinate maps on the torus:

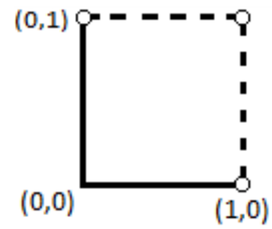


Figure 2.1: A torus.

$$Chart1(x,y) = (x,y) \qquad for\ 0 < x < 1\ and\ 0 < y < 1$$

$$Chart2(x,y) = \begin{cases} (x+\frac{1}{2},y) & : 0 \le x < \frac{1}{2}\ and\ 0 < y < 1 \\ (x-\frac{1}{2},y) & : \frac{1}{2} < x < 1\ and\ 0 < y < 1 \end{cases}$$

$$Chart3(x,y) = \begin{cases} (x,y+\frac{1}{2}) & : 0 < x < 1\ and\ 0 \le y < \frac{1}{2} \\ (x,y-\frac{1}{2}) & : 0 < x < 1\ and\ \frac{1}{2} < y < 1 \end{cases}$$

$$Chart4(x,y) = \begin{cases} (x+\frac{1}{2},y+\frac{1}{2}) & : 0 \le x < \frac{1}{2}\ and\ 0 \le y < \frac{1}{2} \\ (x+\frac{1}{2},y-\frac{1}{2}) & : 0 \le x < \frac{1}{2}\ and\ \frac{1}{2} < y < 1 \\ (x-\frac{1}{2},y+\frac{1}{2}) & : \frac{1}{2} < x < 1\ and\ 0 \le y < \frac{1}{2} \\ (x-\frac{1}{2},y-\frac{1}{2}) & : \frac{1}{2} < x < 1\ and\ \frac{1}{2} < y < 1 \end{cases}$$

These charts clearly cover the whole torus, and it is easy to show that the connections between them are C^∞ (i.e. $Chart1 \circ Chart2^{\leftarrow}$, $Chart2 \circ Chart1^{\leftarrow}$, $Chart1 \circ Chart3^{\leftarrow}$, ...).That satisfies the first two conditions of a differentiable structure on $Torus$. The remaining condition is that our differentiable structure be maximal with respect to those two conditions. This can be done using our atlas of four charts as a sort of "base" to generate the rest:

$$\mathscr{F}_0 = \{Chart1,\ Chart2,\ Chart3,\ Chart4\}$$

$$\mathscr{F} = \{\phi \mid \phi \circ \psi^{\leftarrow} \text{ and } \psi \circ \phi^{\leftarrow} \text{ are } C^\infty \text{ for all } \psi \in \mathscr{F}_0\}$$

One little detail remains. Remember that a manifold is supposed to be a topological space, and its charts are supposed to be homeomorphisms. So how come we never mentioned the topology of $Torus$? And how do we know that the charts are in fact homeomorphisms? The answer is that because we knew how we wanted our charts to map around the torus, we can *define* our topology such that the charts are homeomorphisms. The statement "Take the half open square (figure 2.1) and identify the left and right sides and the top and bottom sides" is closer to what physicists are used to hearing. But we are doing exactly the same thing in a more rigorous way. We are essentially *choosing* our charts such that the generated topology is that of a torus. We define our topology as the result of the following base: The set of inverse images of open sets in the open square in $\mathbb{R}^2$ under any chart in $\mathscr{F}$. That is,

$$\mathcal{B} = \left\{\phi^{\leftarrow}[O] \mid \phi \in \mathscr{F} \bullet O \text{ is an open set in } \mathbb{R}^2\right\}$$

$$\tau = \left\{\bigcup p \mid p \subset \mathcal{B}\right\}$$

where we have explicitly constructed the base and the topology. This concludes the torus example; we have placed a living breathing manifold in front of us.

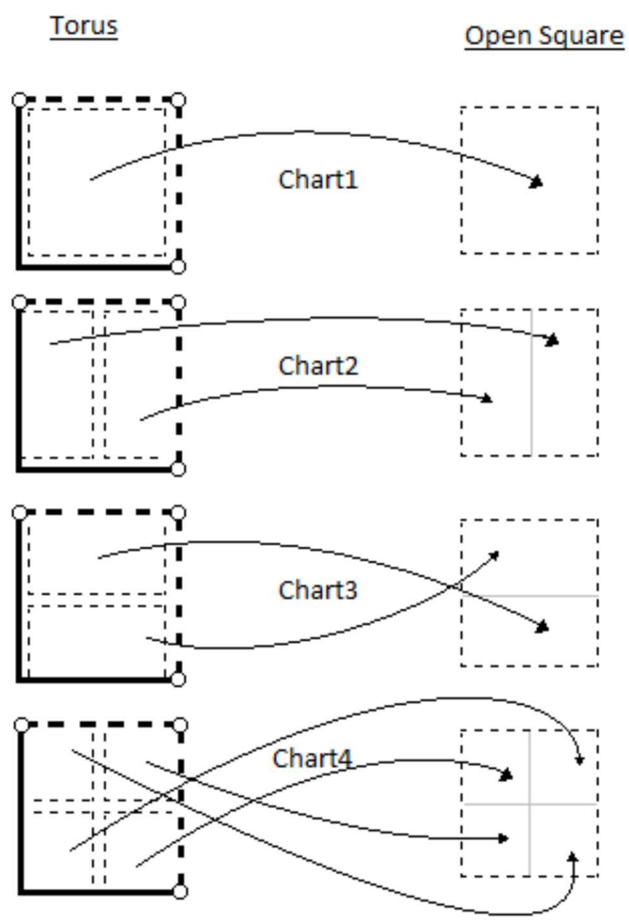


Figure 2.2: A depiction of the four charts.

Chapter 3

Tangent Space

3.1 What we need

When we're doing calculus in regular old $\mathbb{R}^n$, the notion of directional derivative is quite clear. A scalar field, for example, would be a function $f : \mathbb{R}^n \to \mathbb{R}$. Each n-tuple v of real numbers defines a directional derivative operation, $v \cdot \nabla f$, where ∇ represents the primitive gradient operator $(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z})$. The derivative of a path $(\gamma : \mathbb{R} \to \mathbb{R}^n)$ is also pretty simple, it's just the component-wise derivative of the three component functions that make up the path. Then if we want to look at the directional derivative of f along γ, we put the two together: $\gamma' \cdot \nabla f$. Manifolds require us to rethink these ideas.

It is not obvious how one would take a directional derivative of a real valued function on a manifold, or even just the derivative of a path. What we need is a coordinate-independent object that represents direction of travel throughout the manifold. Let us state the problem more precisely. Consider a path in the manifold and a real valued function on the manifold:

$$\gamma : \mathbb{R} \to M$$
$$f : M \to \mathbb{R}$$

How can we define the derivative of f along the path γ? Intuitively such a thing should exist; the space is locally euclidean after all. The composite function $f \circ \gamma$ would give us the values of f along the path according to the path parameter. This is an $\mathbb{R} \to \mathbb{R}^n$ function, so we can take it's derivative at a point and get an n-tuple of real numbers, $(f \circ \gamma)'$. Such a derivative tells us how "fast" the path γ is traveling through values of f. Of course the niceness of f and γ are needed to make this work out, but it seems like the speed at which this path goes through values of f sort of tells us which way it's going and how fast. But the speed we will get obviously depends on choice of f; we need to get the character of f out of our definition, we want the derivative of the path itself. It makes sense to look at the directional derivative given by γ through *any* nice enough function on the manifold. That is, it makes sense to consider the object $(\_\_ \circ \gamma)'$ where the "$\_\_$" can be filled in by any nice enough function.

There is a way of defining tangent vectors that directly uses paths. It actually defines tangent vectors as equivalence classes of paths on the manifold. We will not be doing this. Our definition will take less direct advantage of the path concept, but it will be more powerful. We will be utilizing the algebraic structure of a set of nice functions.

3.2 Definitions

Let $\mathcal{F}_{m,M}$ denote the set of smooth functions on a manifold M that are defined on an open set containing m.

$$\mathcal{F}_{m,M} = \{\ f \mid\ (\exists U \mid f : U \to \mathbb{R} \bullet f \in C^\infty_M \bullet m \in U \in \tau_M)\ \ \}$$

This set seems to have a very natural algebraic structure; it can be made into an algebra. This means it has to have a group operation ("addition"), a multiplication, and a scalar field with a scaling operation. Two elements of this set $f, g \in \mathcal{F}_m$ can be added and multiplied by pointwise addition and multiplication of functions. A real number a can scale a function $f \in \mathcal{F}_m$ in a similar pointwise way.

$$(f + g)(x) = f(x) + g(x)$$

$$(fg)(x) = f(x) \cdot g(x)$$

$$(af)(x) = a \cdot f(x)$$

It is not difficult to show that these operations produce things that are also members of $\mathcal{F}_{m,M}$, nor is it difficult to prove the needed properties of an algebra and more (commutativity, distributivity, associativity, identity, inverse). It should be known that the domain of a function produced after addition or multiplication is the intersection of the domains of the two functions being added or multiplied. (Note: I have a confession to make; this is not really the correct way to construct the algebra. Consider for example the additive inverse of some $f \in \mathcal{F}_{m,M}$, $f : U \to \mathbb{R}$. The inverse should just be $(-1)(f)$ (or $-f$). However upon addition we find that $f + (-1)(f)$ is not just the zero function, it is the zero function restricted to the domain U. For functions to be equal they must have equal domains. For this reason and others, it makes a lot more sense to construct the algebra out of *germs*, equivalence classes of smooth functions that agree on some neighborhood of $m \in M$. Warner provides this treatment in his definition of tangent space, and it is very well done. I am choosing to avoid these complications so as not to lose sight of the basic ideas. It is recommended that the reader who is already familiar with these concepts consult Warner for a fuller explanation).

An *operator* in our case would be a function from $\mathcal{F}_{m,M}$ to the real numbers, $\mathcal{O} : \mathcal{F}_{m,M} \to \mathbb{R}$. A *linear* operator has the property $\mathcal{O}(af + bg) = a \cdot \mathcal{O}(f) + b \cdot \mathcal{O}(g)$ for any $f, g \in \mathcal{F}_{m,M}$, $a, b \in \mathbb{R}$. An operator which is a *derivation-at-m* has the property $\mathcal{O}(fg) = \mathcal{O}(f) \cdot g(m) + \mathcal{O}(g) \cdot f(m)$ for any $f, g \in \mathcal{F}_{m,M}$. Operators also have their own pointwise addition and scalar multiplication over the real numbers. For two operators $\mathcal{O}, \mathcal{P}$ and $a \in \mathbb{R}$

$$(\mathcal{O} + \mathcal{P})(f) = \mathcal{O}(f) + \mathcal{P}(f)$$

$$(a\mathcal{O})(f) = a \cdot \mathcal{O}(f)$$

Definition 3.2.1. *The* ***tangent space*** *of the manifold M at $m \in M$, $\mathcal{T}_{m,M}$, is the set of linear derivations-at-m on $\mathcal{F}_{m,M}$. This is a vector space over the real numbers, with addition and scalar multiplication being the above defined operator operations.*

Let us now make sure that our space behaves in the way we intended. If $a \in \mathbb{R}$, let $\bar{a}$ denote the constant function of value a on the manifold. Clearly $\bar{a} \in \mathcal{F}_{m,M}$ for any $m \in M$.

Theorem 3.2.1. *For any $t \in \mathcal{T}_{m,M}$ and $c \in \mathbb{R}$, $t(\bar{c}) = 0$.*

Proof. $c \cdot t(\bar{1}) = t(\bar{c} \cdot \bar{1}) = t(\bar{c}) \cdot \bar{1}(m) + t(\bar{1}) \cdot \bar{c}(m) = t(\bar{c}) \cdot 1 + t(\bar{1}) \cdot c = t(\bar{c}) + c \cdot t(\bar{1})$
So $c \cdot t(\bar{1}) = t(\bar{c}) + c \cdot t(\bar{1})$ so $t(\bar{c}) = 0$ □

Theorem 3.2.2. *For any $f, g \in \mathcal{F}_{m,M}$ such that f and g have the same value on some neighborhood U of m, we have $t(f) = t(g)$ for any $t \in T_{m,M}$.*

Proof. Take $f, g \in \mathcal{F}_m$ such that f and g have the same value on the neighborhood U of m, and take any $t \in \mathcal{T}_{m,M}$. Let $\bar{1}_{\restriction U}$ denote the function $\bar{1}$ restricted to the domain U. Note that our hypothesis on f and g can be written $(\bar{1}_{\restriction U}) \cdot f = (\bar{1}_{\restriction U}) \cdot g$.
$t((\bar{1}_{\restriction U}) \cdot f) = t(\bar{1}_{\restriction U}) \cdot f(m) + t(f) \cdot (\bar{1}_{\restriction U})(m) = t(\bar{1}_{\restriction U}) \cdot f(m) + t(f) = t(\bar{1}_{\restriction U} \cdot g) = t(\bar{1}_{\restriction U}) \cdot g(m) + t(g) = t(\bar{1}_{\restriction U}) \cdot f(m) + t(g)$
where we have used $f(m) = g(m)$. We established $t(\bar{1}_{\restriction U}) \cdot f(m) + t(f) = t(\bar{1}_{\restriction U}) \cdot f(m) + t(g)$, so $t(f) = t(g)$. □

This basically says that a tangent vector at m only cares about purely local values of functions around m. If it wasn't clear already, this theorem should make it clear that it makes a lot more sense to deal with an algebra consisting of germs of functions rather than actual functions. Nevertheless, we are happy to see tangent vectors work the way we wanted them to.

Now we may answer the original question. If $\gamma : \mathbb{R} \to M$ is a smooth curve and $\gamma(c) = m$ then we define the tangent vector to γ at c to be the operator t such that for $f \in \mathcal{F}_{m,M}$,

$$t(f) = (f \circ \gamma)'(c)$$

We characterize the speed and direction at which a path is traveling through the manifold by the speed at which it travels through the values of smooth real-valued functions on the manifold. It is not difficult to show that the definition above is in fact a linear derivation at m, and thus may be called a tangent vector at m.

3.3 Coordinate Basis

Now let us consider the dimensionality of $\mathcal{T}_{m,M}$ for some manifold. Wald's *General Relativity* (p17) does a nice proof that the dimensionality of the tangent space is the same as the dimensionality of the manifold. I will not show this proof here, but the important part is that he does it by constructing a basis for $T_{m,M}$. I recommend Bishop and Goldberg (p52) for proofs about the basis. First, get a chart ϕ that has m in its domain (one should exist since the charts cover the manifold). Let $\{t_i\}$ be tangent vectors such that for any $f \in \mathcal{F}_{m,M}$,

$$t_i(f) = \left.\frac{\partial f \circ \phi^{\leftarrow}}{\partial x^i}\right|_{\phi(m)}$$

Since f maps from M to $\mathbb{R}$ and $\phi^{\leftarrow}$ maps from $\mathbb{R}^n$ to M, the object $f \circ \phi^{\leftarrow}$ is from $\mathbb{R}^n$ to $\mathbb{R}$. It is just the function f viewed in $\mathbb{R}^n$ through the eyes of the coordinate system ϕ. And the t_i are the operators that look at functions through ϕ and take their derivatives in the directions of the coordinate axes. They form a very special basis called the *coordinate basis*. That they form a basis at all is an important theorem whose proof I will not show, but it is in the references mentioned above. It is an excellent exercise to show, using the definitions we've provided, that the coordinate basis at a point consists of the tangent vectors to the coordinate axes (coordinate axes are actual paths in the manifold).

We conclude this section by examining a property of tangent vectors that many physicists use to *define* the word "vector". This is the transformation law that arises from the coordinate basis expansion of an arbitrary $t \in \mathcal{T}_{m,M}$, and the chain rule. I like to express operators by using an underscore to show how they behave. For example the coordinate basis vector at m for the i^{th} coordinate of a coordinate system ϕ is a map that takes $f \in \mathcal{F}_{m,M}$ and spits out the number $\frac{\partial(f\circ\phi^{\leftarrow})}{\partial x^i}\Big|_{\phi(m)}$. My informal way of denoting this operation is $\frac{\partial(\_\_\circ\phi^{\leftarrow})}{\partial x^i}\Big|_{\phi(m)}$. The underscore looks like a blank space where the operator is waiting to be fed a function in $\mathcal{F}_{m,M}$. Now take any $t \in \mathcal{T}_{m,M}$. Let's express it in the coordinate basis induced by the coordinate system ϕ, and then try to get its expression in terms of the coordinate system ψ.

$$
\begin{aligned}
t &= \sum_{i=1}^{n} a^i \frac{\partial(\_\_\circ\phi^{\leftarrow})}{\partial x^i}\Bigg|_{\phi(m)} \\
&= \sum_{i=1}^{n} a^i \frac{\partial(\_\_\circ\psi^{\leftarrow}\circ\psi\circ\phi^{\leftarrow})}{\partial x^i}\Bigg|_{\phi(m)} \\
&= \sum_{i=1}^{n} a^i \sum_{j=1}^{n} \frac{\partial(\_\_\circ\psi^{\leftarrow})}{\partial x^j}\Bigg|_{\psi(\phi^{\leftarrow}(\phi(m)))} \frac{\partial(\psi\circ\phi^{\leftarrow})^j}{\partial x^i}\Bigg|_{\phi(m)} \\
&= \sum_{j=1}^{n} \left(\sum_{i=1}^{n} a^i \frac{\partial(\psi\circ\phi^{\leftarrow})^j}{\partial x^i}\Bigg|_{\phi(m)} \right) \frac{\partial(\_\_\circ\psi^{\leftarrow})}{\partial x^j}\Bigg|_{\psi(m)}
\end{aligned}
$$

The superscript in $(\psi\circ\phi^{\leftarrow})^j$ indicates the j^{th} component function. So the new components in the ψ coordinate basis depend on the coordinate transformation $(\psi\circ\phi^{\leftarrow})$.

$$
b^j = \sum_{i=1}^{n} a^i \frac{\partial(\psi\circ\phi^{\leftarrow})^j}{\partial x^i}\Bigg|_{\phi(m)}
$$

$$
t = \sum_{i=1}^{n} b^i \frac{\partial(\_\_\circ\psi^{\leftarrow})}{\partial x^i}\Bigg|_{\psi(m)}
$$

This is the tangent vector transformation law.

Chapter 4

Cotangent Space

I read what four books had to say about this topic since it's important to understand before getting into differential forms and tensors. Warner was the most general and the most concise, but I eventually chose to follow Frankel's approach. Schutz spent a lot of time on fiber bundles and tangent bundles in particular before getting into differential forms. His presentation of bundles was extremely good, but since my goal in these notes is to get to lie derivatives, I'm not doing bundles. Frankel doesn't assume the reader's familiarity with dual spaces. He takes care of the necessary math with good explanations, and then he defines differentials with the tools he built. The only problem is that he doesn't treat the more general definition first. So I'm going to start by presenting the general, formal definition of a differential up front. Then I'll backtrack to the beginning and build up to the specific kind of differential we'll be using. For learning this the first time, Bishop and Goldberg chapter 2 was the best reading.

Take note that this discussion is limited to finite dimensional vector spaces. I don't need the infinite dimensional results, and there are some serious differences. Also, the scalar field over which these vector spaces lie is the field of real numbers. I didn't have to do this, but it's all we really need. It's easy to make the discussion general; just replace the word "real number" with "scalar" and replace $\mathbb{R}$ with F.

4.1 A Cold Definition

Consider a C^∞ map $\psi : M \to N$ from one manifold to another. The differential of ψ at $m \in M$ is a linear map $d\psi : \mathcal{T}_{m,M} \to \mathcal{T}_{\psi(m),N}$ defined as follows. For any tangent vector $t \in \mathcal{T}_{m,M}$, feeding it to the differential should give a tangent vector $d\psi(t) \in \mathcal{T}_{\psi(m),N}$ in the tangent space of the other manifold. This tangent vector $d\psi(t)$ is defined by how it acts on some $g \in \mathcal{F}_{\psi(m),N}$:

$$d\psi(t)(g) = t(g \circ \psi)$$

So the map ψ is used to view the function on N, g, as it appears on M through ψ's eyes. This is then what is fed to t. The notation $d\psi$ is actually incomplete; a good notation should indicate information about m and M (for example $d\psi_{m,M}$). But we'll stick to crappy notation for now. Well that's it for the definition, it doesn't really tell us much. We will come back to this definition more seriously in section 7.1. For now what we will be interested in is a special case of this definition where the second manifold, N, is actually just $\mathbb{R}$. Let us go back to the beginning and make some sense of this.

4.2 Dual Space

Consider a vector space E over the real numbers. A *linear function* on a vector space E is a function $h : E \to \mathbb{R}$ such that for $a, b \in \mathbb{R}$ and $v, w \in E$, $h(av + bw) = ah(v) + bh(w)$. The property of linearity is a pretty big deal here. Since it generalizes to any finite sum by induction, and we can expand any vector $v \in E$ into components of a basis $\{\hat{e}_j\}$,

$$h(v) = h\left(\sum_{j=1}^{n} \hat{e}_j v^j\right) = \sum_{j=1}^{n} h(\hat{e}_j) v^j$$

we see that the action of a linear function on a *any* vector in E is completely defined by what it does to only the vectors of a basis of E. This means we can just take any old basis $\{\hat{e}_j\}$ of E, say what number each $h(\hat{e}_j)$ gives, and define h to be the "linear extension" of that.

So where do these linear functions live? They live in another vector space E^* called the *dual space* of E. E^* is just the set of all linear functions on E. Its elements are often called dual vectors. It has the pointwise vector addition and scalar multiplication of functions. So for $a \in \mathbb{R}$, $w_1, w_2 \in E^*$, and any $v \in E$:

$$(a \cdot w_1)(v) = a \cdot w_1(v)$$

$$(w_1 + w_2)(v) = w_1(v) + w_2(v)$$

It's easy to show that the dual space is a vector space. Now for any basis of E, $\{\hat{e_i}\}$, there is a corresponding basis of E^*, $\{\hat{\sigma}^i\}$, defined by the linear extensions of:

$$\hat{\sigma}^i(\hat{e}_j) = \delta^i_j$$

It is not difficult to show that these form a basis for E^*. Something interesting happens when we feed a vector $v \in E$ to a basis dual vector $\hat{\sigma}^i$.

$$\hat{\sigma}^i(v) = \hat{\sigma}^i\left(\sum_{j=1}^{n} \hat{e}_j v^j\right) = \sum_{j=1}^{n} \hat{\sigma}^i(\hat{e}_j) v^j = \sum_{j=1}^{n} \delta^i_j v^j = v^i$$

So if $\hat{\sigma}^i$ is the dual basis vector corresponding to a basis vector $\hat{e}_i$, then feeding any $v \in E$ to $\hat{\sigma}^i$ gives us v's component along $\hat{e}_i$. Now we can find the basis expansion of an arbitrary dual vector $w \in E^*$ by just feeding it a vector $v \in E$ and using what we just found:

$$w(v) = w\left(\sum_{j=1}^{n} v^j \hat{e}_j\right) = \sum_{j=1}^{n} v^j w(\hat{e}_j) = \sum_{j=1}^{n} \hat{\sigma}^j(v) w(\hat{e}_j)$$

$$w = \sum_{j=1}^{n} w(\hat{e}_j)\ \hat{\sigma}^j$$

This means that the component of w along $\hat{\sigma}^j$ is $w(\hat{e}_j)$.

When we talk about a $w \in E^*$, we're talking about a real valued function of an E-valued variable, v.

$$w(v), \quad \text{for } v \in E$$

We can switch around our view of the real number $w(v)$. Instead of focusing our attention on this particular $w \in E^*$, we can consider a particular $v \in E$. We can then see $w(v)$ as a real valued function of an E^*-valued variable, w.

$$w(v), \quad \text{for } w \in E^*$$

So what we're really talking about is a $\bar{v} : E^* \to \mathbb{R}$ such that for any $w \in E^*$,

$$\bar{v}(w) = w(v)$$

We see that $\bar{v} \in E^{**}$, it is a member of the dual space to the dual space of E. The map $v \mapsto \bar{v}$ can be shown to be a 1-1 linear function from E to E^{**}. This makes the dual space of the dual space of a vector space isomorphic to the original vector space. This is no ordinary vector space isomorphism, however. This is a *natural* isomorphism. The real definition of natural isomorphism comes from category theory; the actual natural isomorphism is the functor one can define from the entire category of n dimensional vector spaces to itself. There is no real need to get into the details of category theory, so can just be thought of as a vector space isomorphism that is related only to the vector space structure, not any basis. For example, the dual space to a vector space is isomorphic to the vector space, but this isomorphism is not natural because it depends on choice of basis. Any n-dimensional vector space is isomorphic to $\mathbb{R}^n$, but again there is an isomorphism for each choice of basis.

We could have also viewed the object $w(v)$ in the following symmetric way: $w(v)$ is a real valued function of two variables, an E-valued variable v and an E^*-valued variable w. The map $(v, w) \mapsto w(v)$ is then our definition of scalar product: $< v, w >$.

4.3 Cotangent Space

The tangent space at a point in a manifold, $\mathcal{T}_{m,M}$ is a vector space. The dual of this space, $\mathcal{T}^*_{m,M}$, is called the *cotangent space*. We can define an element of the cotangent space at m by using a function $f \in \mathcal{F}_{m,M}$. It is called the *differential of f*, $df : \mathcal{T}_{m,M} \to \mathbb{R}$. For any $v \in \mathcal{T}_{m,M}$, it is defined by

$$df(v) = v(f)$$

So taking a smooth real valued function f on the manifold and putting a d in front of it creates a function that feeds f to any tangent vector fed to it. This is obviously a linear function on $\mathcal{T}_{m,M}$, so it must be in $\mathcal{T}^*_{m,M}$. I find it helpful to resort to my underscore notation, which I explained in the previous section and in the appendix. In terms of the components of a tangent vector in a coordinate basis from a coordinate system ϕ,

$$df(v) = df\left(\sum_{j=1}^{n} v^j \frac{\partial \_\_ \circ \phi^{\leftarrow}}{\partial x^j}\bigg|_{\phi(m)}\right) = \sum_{j=1}^{n} v^j \frac{\partial f \circ \phi^{\leftarrow}}{\partial x^j}\bigg|_{\phi(m)}$$

It's nice to look at tangent vectors and their duals in terms of the underscore notation, in a coordinate basis. A vector, given components v^j, is just the map

$$\sum_{j=1}^{n} v^j \frac{\partial \_\_ \circ \phi^{\leftarrow}}{\partial x^j}\bigg|_{\phi(m)}$$

where the underscore is waiting to be fed a smooth function. And a dual vector, given a smooth function f, is the map

$$\sum_{j=1}^{n} \_\_^j \frac{\partial f \circ \phi^{\leftarrow}}{\partial x^j}\bigg|_{\phi(m)}$$

where the underscore is waiting to be fed the components of a vector in the coordinate basis. This view makes it intuitively clear that the dual of the dual to the tangent space is naturally isomorphic to the tangent space.

Let us now consider the differential of a coordinate function. Suppose $(x^1, ..., x^n)$ are the coordinate functions (component functions, projections) of a coordinate system ϕ. First look at how dx^i acts on the j^{th} coordinate basis vector:

$$dx^i \left(\left. \frac{\partial_{--} \circ \phi^{\leftarrow}}{\partial x^j} \right|_{\phi(m)} \right) = \left. \frac{\partial x^i \circ \phi^{\leftarrow}}{\partial x^j} \right|_{\phi(m)} = \delta^i_j$$

Now we're able to decompose the action of the differential on any tangent vector:

$$dx^i \left(\sum_{j=1}^{n} v^j \left. \frac{\partial_{--} \circ \phi^{\leftarrow}}{\partial x^j} \right|_{\phi(m)} \right) = \sum_{j=1}^{n} v^j \left. \frac{\partial x^i \circ \phi^{\leftarrow}}{\partial x^j} \right|_{\phi(m)} = \sum_{j=1}^{n} v^j \delta^i_j = v^i$$

The differential of the i^{th} coordinate function just reads off the i^{th} component of a tangent vector in the coordinate basis! The $\{dx^i\}$ then form a dual basis to the coordinate basis. That is, we can express any linear function in terms of this dual basis expansion:

$$w = \sum_{j=1}^{n} w \left(\frac{\partial_{--} \circ \phi^{\leftarrow}}{\partial x^j} \right) dx^j$$

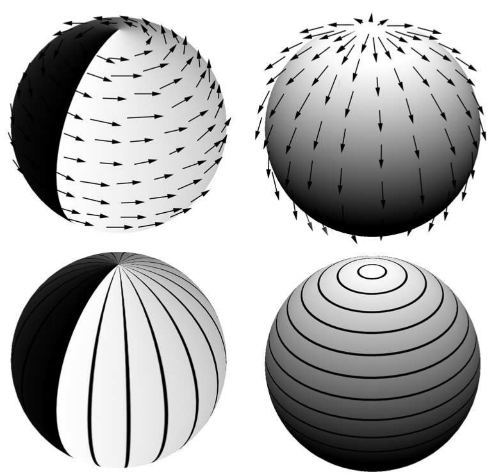


Figure 4.1: For the two-sphere as a manifold, this is an artistic representation of coordinate-basis vectors (top) and dual vectors (bottom), for the typical theta-phi spherical coordinate system. The shading represents the value of the coordinate functions. We may visualize a vector t as an arrow pointing in the direction that a smooth function f would have to change so that $t(f)$ is a big real number. We may visualize dual vectors as level curves of a function f around a point, level curves that would have to punctured by a vector t so that $df(t)$ can be a big real number. See if you can pictorially convince yourself that the dual basis pictures on the bottom each correspond to the basis vectors above them. But don't make too much of the picture.

Chapter 5

Tensors

This section will begin with more talk of linear functions on vector spaces. We will discuss matrix forms, multilinear functions, summation notation, tensor spaces, tensor algebra, natural isomorphisms leading to different interpretations of tensors, bases of tensor spaces and components of tensors, and transformation laws. There will be a tiny bit of repetition of the previous section, because this is its ultimate generalization. I learned most of this material from Bishop and Goldberg chapter 2, which I highly recommend. For someone who is very familiar with algebra (not me), I would recommend Warner. Schutz gives the typical physics explanation of tensors, which is a very bad explanation for a newcomer to the subject. Like the previous section, I'm going to start with general mathematical concepts before applying them to our concrete example of a vector space, the tangent space. For my own reference, here is a list of topics I missed that I would like to fill in someday: invariants, formal definition of a contraction, symmetric algebra, grassman algebra, determinant and trace, and hodge duality.

5.1 Linear Functions and Matrices

Let us expand our previous definition of linear function. A linear function $f : V \to W$ may now map from one vector space V into another W so that for $v_1, v_2 \in V$ and $a \in \mathbb{R}$ we have

$$f(v_1 + v_2) = f(v_1) + f(v_2)$$

$$f(av_1) = af(v_1)$$

Bijective linear functions are vector space isomorphisms. The set of linear functions from V to W forms a vector space with pointwise addition of functions and pointwise scalar multiplication. There are a lot of nice things we could prove about bases, dimensionality, null spaces, and image spaces. For this I refer the reader to Bishop and Goldberg pages 59-74, or a linear algebra book. I don't have time to get that material written but we'll build up what we need.

It is still the case that we only have to define linear functions on basis elements to characterize them. Consider a linear function $f : V \to W$, where $\dim(V)=d_V$ and $\dim(W)=d_W$. Let $\{\hat{v}_i\}$ be a basis of V and $\{\hat{w}_\alpha\}$ a basis of W, where the indices run over the appropriate dimension. We introduce the hopefully familiar summation convention, in which any index that appears twice is summed over. Let's say that we've determined the action of f on the basis vectors $\{\hat{v}_i\}$. Since f takes each of those to a particular vector in W, we will need d_W different scalars to represent each $f(\hat{v}_i)$ in the basis $\{\hat{w}_\alpha\}$,

$$f(\hat{v}_i) = a_i^\alpha \hat{w}_\alpha$$

The double-indexed object a_i^α is a d_W by d_V matrix, where α is the row index and i is the column index. a is called the matrix of f with respect to $\{\hat{v}_i\}$ and $\{\hat{w}_\alpha\}$. Notice that our designation of one index as "row" and the other as "column" is completely arbitrary. This is the point in our discussion where we make that arbitrary choice, and any other remark about rows or columns will be with respect to this choice. Now consider any $v \in V$. Then $f(v)$ is some $w \in W$.

$$f(v) = f(v^i \hat{v}_i) = v^i f(\hat{v}_i) = v^i a_i^\alpha \hat{w}_\alpha = w^\alpha \hat{w}_\alpha$$

$$\text{where } w^\alpha = a_i^\alpha v^i$$

The action of f is in fact determined by our grid of numbers a_i^α. We let this give us the definition of multiplication of a matrix by what we now deem to be a column matrix v^i to give another column matrix w^i. Let $B_V : V \to \mathbb{R}^{d_V}$ be the function that takes any vector $v \in V$ and gives the d_V-tuple of components in the basis $\{\hat{v}_i\}$, and similarly for $B_W : W \to \mathbb{R}^{d_W}$. Let $A : \mathbb{R}^{d_V} \to \mathbb{R}^{d_W}$ represent matrix multiplication of a_i^α by the tuples in $\mathbb{R}^{d_V}$. The following diagram then sums up what we just showed:

$$\begin{tikzcd} V \arrow[r, "f"] \arrow[d, "B_V"'] & W \arrow[d, "B_W"] \\ \mathbb{R}^{d_V} \arrow[r, "A"'] & \mathbb{R}^{d_W} \end{tikzcd}$$

We could define matrix multiplication in terms of the composition of operators, and we could prove distributivity and associativity. I will not do this here.

Summation notation is pretty tricky. It should be noted that the types of sums being done depends on the context. For example $v^i \hat{e}_i$ is a sum of vectors, vector addition. But $v^i w_i$ is a sum of scalars, real number addition. Summation notation makes everyday tasks easy, but it also obscures some other things. For example, what do we really mean by "a_i^α?" Is it a matrix? Or is it a number, the $(\alpha, i)^{th}$ component of a matrix? Or is it the operator independent of any basis? Authors don't seem to agree on the answer to this question. Wald solves the problem by using latin indices for operators and greek indices for components in a basis. Here is my system: a is the operator, the function on the vector space(s). $\mathbf{a}_i^\alpha$ is still the operator, but with indices that indicate the structure of the operator if it were to be resolved in a basis. Contractions of indices in this case are basis independent things done to operators. a_i^α is the $(\alpha, i)^{th}$ component of a matrix in some basis, and contractions of these indices are actual sums.

5.2 Multilinear Functions

A function $f : V_1 \times V_2 \to W$ is *multilinear* if it is linear in each individual variable. That is, for $v_1, y_1 \in V_1$, $v_2, y_2 \in V_2$, and $a, b \in \mathbb{R}$

$$f(av_1 + by_1, v_2) = af(v_1, v_2) + bf(y_1, v_2)$$

$$f(v_1, av_2 + by_2) = af(v_1, v_2) + bf(v_1, y_2)$$

The function above is bilinear. The definition of multilinear generalizes to a function of n vectors,

$$f(v_1, \dots, av_i + by_i, \dots, v_n) = af(v_1, \dots, v_i, \dots, v_n) + bf(v_1, \dots, y_i, \dots, v_n)$$

Now suppose $\nu \in V^*$ and $\omega \in W^*$, these are linear real-valued functions on the vector spaces V and W. We can form a bilinear function on $V \times W$ by taking their *tensor product*. It is a function such that for $v \in V$ and $w \in W$

$$\nu \otimes \omega(v, w) = \nu(v) \cdot \omega(w)$$

We can pointwise-add multilinear functions of a certain type to produce more multilinear functions. We can also scalar multiply multilinear functions pointwise. Thus the set of multilinear functions on some vector spaces $V_1, V_2, \dots, V_n$ into W is a vector space. We denote this vector space by $L(V_1, \dots, V_n; W)$.

5.3 Tensors

For a vector space V, the real-valued multilinear functions with any number of variables in V and V^* are called *tensors* over V. The tensor type is determined by the number of dual vectors it takes and the number of vectors it takes, in that order. A multilinear function $T : V^* \times V \times V \to \mathbb{R}$ for example is a type $(1, 2)$ tensor. We always want to have the dual spaces come first in the list of variables, so we are not interested in a map $V \times V^* \to \mathbb{R}$. That map is already taken care of by the equivalent one with permuted variables, $V^* \times V \to \mathbb{R}$. The set of tensors of a particular type form a vector space called a *tensor space over* V, and we call that space $T^i_j(V)$ for tensors of type (i, j) over V. Addition and scalar multiplication are pointwise for functions. So for $A, B \in T^i_j(V)$, $a \in \mathbb{R}$, $v_1, \dots v_j \in V$, and $w_1, \dots, w_i \in V^*$

$$(A + B)(w_1, \dots, w_i, v_1, \dots, v_j) = A(w_1, \dots, w_i, v_1, \dots, v_j) + B(w_1, \dots, w_i, v_1, \dots, v_j)$$

$$(aA)(w_1, \dots, w_i, v_1, \dots, v_j) = aA(w_1, \dots, w_i, v_1, \dots, v_j)$$

A tensor of type $(0, 0)$ is defined to be a scalar, $T^0_0(V) = \mathbb{R}$. Notice that we are treating the vector space V as if it *is* V^{**}. The natural isomorphism we talked about is taken very seriously. People hop between V and V^{**} so effortlessly that we just use V to refer to both.

So $T^i_j(V)$ is the set of multilinear real-valued functions on V^* i times and on V j times. In particular, $T^1_1(V)$ is the set of multilinear maps on $V^* \times V$. If we take a $v_1 \in V$ and a $w_1 \in V^*$, we can form an element of $T^1_1(V)$ using the tensor product we defined earlier, $v_1 \otimes w_1 \in T^1_1(V)$. Can we form any element of $T^1_1(V)$ in this way? The answer is no, if the dimension of V is at least 2. Because for dimension 2 for example, we could take another linearly independent $v_2 \in V$ and $w_2 \in V^*$. Then if we were able to formulate $v_1 \otimes w_1 + v_2 \otimes w_2$ as some $\bar{v} \otimes \bar{w}$, we could partially evaluate the resulting equation, $v_1 \otimes w_1 + v_2 \otimes w_2 = \bar{v} \otimes \bar{w}$, and we would be forced to violate linear independence. We will soon see how we can form a basis for tensor spaces, but let's get some algebra out of the way first.

We should expand our first definition of tensor product to work for tensors of any type. The tensor product of an (i, j) tensor A with an (r, s) tensor B is an $(i + r, j + s)$ tensor that takes in the combined variables of A and B, feeds them respectively to A and B to get two real numbers, and then spits out the product. That is, for a set of $w^a \in V^*$ and $v_b \in V$

$$(A \otimes B)(w^1, \dots, w^{i+r}, v_1, \dots, v_{j+s}) = A(w^1, \dots, w^i, v_1, \dots, v_j) \cdot B(w^{i+1}, \dots, w^{i+r}, v_{j+1}, \dots, v_{j+s})$$

Now we can state associative and distributive laws for tensor product, they are easy to prove:

$$(A \otimes B) \otimes C = A \otimes (B \otimes C)$$

$$A \otimes (B + C) = A \otimes B + A \otimes C$$
$$(A + B) \otimes C = A \otimes C + B \otimes C$$

We already have some examples of tensor spaces. A vector space V^{**} (which we identify with V) contains linear functions on V^*, so it is just the tensor space $T^1_0(V)$. The dual vector space contains linear functions on V, so it is just the tensor space $T^0_1(V)$.

To resolve the elements of a tensor space $T^r_s(V)$ in a particular basis, we need a basis of V, $\{\hat{e}_i\}$, and a basis of V^*, $\{\hat{\sigma}^j\}$. Just like a linear function is completely determined by its action on a basis and its linear extension, a multilinear function is determined by its action on the basis $\{\hat{e}_{i_1} \otimes \ldots \otimes \hat{e}_{i_r} \otimes \hat{\sigma}^{j_1} \otimes \ldots \otimes \hat{\sigma}^{j_s}\}$ (for all permutations of the indices) and its multilinear extension. If we can prove that the action of $A \in T^r_s(V)$ on $\{\hat{e}_{i_1} \otimes \ldots \otimes \hat{e}_{i_r} \otimes \hat{\sigma}^{j_1} \otimes \ldots \otimes \hat{\sigma}^{j_s}\}$ completely determines A, then we can express A in terms of the proposed basis and we know that it spans the space. Linear independence is easy to prove using induction. So we can show that $\{\hat{e}_{i_1} \otimes \ldots \otimes \hat{e}_{i_r} \otimes \hat{\sigma}^{j_1} \otimes \ldots \otimes \hat{\sigma}^{j_s}\}$ is a basis for $T^r_s(V)$. Let

$$A^{i_1 \ldots i_r}_{j_1 \ldots j_s} = A(\hat{\sigma}^{i_1}, \ldots, \hat{\sigma}^{i_r}, \hat{e}_{j_1}, \ldots, \hat{e}_{j_s})$$

To feed A an arbitrary set of variables we need a set of r dual vectors $\omega^a \in V^*$ and a set of s vectors $\nu_b \in V$. These can expressed in their components in each basis:

$$\omega^a = w^a_k \hat{\sigma}^k$$

$$\nu_b = v^l_b \hat{e}_l$$

And just because we'll need it soon, remember that basis vectors and dual vectors can be used to read off components of dual vectors and vectors:

$$w^a_k = \hat{e}_k(\omega^a)$$
$$v^l_b = \hat{\sigma}^l(\nu_b)$$

Now we can feed them to A,

$$A(\omega^1, \ldots, \omega^r, \nu_1, \ldots, \nu_s) = A(w^1_{i_1}\hat{\sigma}^{i_1}, \ldots, w^r_{i_r}\hat{\sigma}^{i_r}, v^{j_1}_1 \hat{e}_{j_1}, \ldots, v^{j_s}_s \hat{e}_{j_s})$$

and use multilinearity:

$$\begin{aligned}
&= (w^1_{i_1} \ldots w^r_{i_r})(v^{j_1}_1 \ldots v^{j_s}_s) A(\hat{\sigma}^{i_1}, \ldots, \hat{\sigma}^{i_r}, \hat{e}_{j_1}, \ldots, \hat{e}_{j_s}) \\
&= (w^1_{i_1} \ldots w^r_{i_r})(v^{j_1}_1 \ldots v^{j_s}_s) A^{i_1 \ldots i_r}_{j_1 \ldots j_s} \\
&= A^{i_1 \ldots i_r}_{j_1 \ldots j_s} (\hat{e}_{i_1}(\omega^1) \ldots \hat{e}_{i_r}(\omega^r))(\hat{\sigma}^{j_1}(\nu_1) \ldots \hat{\sigma}^{j_s}(\nu_s)) \\
&= \left[A^{i_1 \ldots i_r}_{j_1 \ldots j_s} (\hat{e}_{i_1} \otimes \ldots \otimes \hat{e}_{i_r} \otimes \hat{\sigma}^{j_1} \otimes \ldots \otimes \hat{\sigma}^{j_s})\right] (\omega^1, \ldots, \omega^r, \nu_1, \ldots, \nu_s)
\end{aligned}$$

And we finally have A in terms of the basis:

$$A = A^{i_1 \ldots i_r}_{j_1 \ldots j_s} (\hat{e}_{i_1} \otimes \ldots \otimes \hat{e}_{i_r} \otimes \hat{\sigma}^{j_1} \otimes \ldots \otimes \hat{\sigma}^{j_s})$$

Notice that the dimension of the tensor space $T^r_s(V)$ must then be $\dim(V)^{r+s}$

5.4 Interpretations of Tensors

Consider a tensor $A \in T^1_1(V)$. There are a couple of ways to view this object. One could see it for what it is, a multilinear map from $V^* \times V$ to the real numbers. This isn't always the geometrically meaningful interpretation. If we take a dual vector $w \in V^*$ and we feed that to A, then we've only partially evaluated it. A still has its mouth open, waiting to be fed a vector. So after feeding only w to A we're left with a linear map $V \to \mathbb{R}$, which is a dual vector. We can name this function $A_1 : V^* \to V^*$, and it is the mapping $w \mapsto (v \mapsto A(w,v))$. Similarly we could have taken a $v \in V$, fed v to A, and left it hungry for a dual vector. That would have left us with a map $V^* \to \mathbb{R}$, which is a vector (in the naturally isomorphic sense). We name that function $A_2 : V \to V$ and it is the mapping $v \mapsto (w \mapsto A(w,v))$. These are three ways to interpret the same action: producing a real number from two objects. We consider the different mappings A, A_1, and A_2 to be the same without any choice of basis; so they are *naturally* the same. Formally we would say that $T^1_1(V),\ L(V;V),$ and $L(V^*;V^*)$ are naturally isomorphic vector spaces.

These interpretations translate over to different designations of rows and columns once A is resolved in some basis $\left\{\hat{e}_i \otimes \hat{\sigma}^j\right\}$. Higher rank tensors analogously have multiple interpretations, there are more kinds of interpretations for higher ranks. In practical situations, people treat naturally isomorphic spaces as the same space without making a fuss over it. Even though different interpretations of the same tensor are strictly different functions, people mix all of the different mappings into one "index structure". So when we say $\mathbf{g}_{\mu\nu}$ we are referring to all of the interpretations mushed into one object. Once indices are contracted in a particular way, like $\mathbf{g}_{\mu\nu}\mathbf{x}^\mu$, we pick one interpretation (in this case the $V \to V^*$ type).

I'd like to do an example of a tensor interpretation with scalar products, before I end this. The scalar product operation, mentioned at the end of section 4.2, takes a vector and a dual vector and gives back a real number. It does this in a bilinear way, which makes it a tensor $<,>: V^* \times V \to \mathbb{R}$. That's one interpretation, $<,> \in T^1_1(V)$. What about $<,>_1: V^* \to V^*$? Well if we feed the scalar product a dual vector $w \in V^*$, it's still waiting to be fed a vector. The map we are left with wants to take the vector we give it and feed it to w, by definition of scalar product. The remaining map must then be w itself. Okay how about $<,>_2: V \to V$? Feed $<,>$ a vector $v \in V$, and it is left hungry for a dual vector. When the remaining map is fed a dual vector, it just feeds that to v and gives us the result. The remaining map must then be v itself by definition. So $<,>_1$ and $<,>_2$ are identity maps. The components of $<,>$ in a basis $\left\{\hat{e}_i \otimes \hat{\sigma}^j\right\}$ are $<,>^i_j = <\hat{\sigma}^i, \hat{e}_j> = \delta^i_j$. It's a Kronecker delta in any basis.

We could talk about tensor transformation laws in terms of vector transformation laws completely abstractly, without using our tangent space. But it's time we head back home to the manifold; just know that the results concerning transformation laws for tensors over a tangent space are derivable in general. It's more instructive and practical to get the transformation laws through the concrete example of tensor spaces over a tangent space.

5.5 Back on the Manifold

The vector space we've been interested in is the tangent space at a point m in a manifold M, $\mathcal{T}_{m,M}$. The tensor spaces we are interested in are then $T^r_s(\mathcal{T}_{m,M})$. Let ϕ, ψ be coordinate systems with $m \in M$ in their domain. Let the component functions of ϕ be $x^1, \ldots, x^n : M \to \mathbb{R}$, where n is the dimension of the manifold. Let the component functions of ψ be $y^1, \ldots, y^n$. Let $t \in \mathcal{T}_{m,M}$ be a tangent vector at m, and let ${}^\phi t^i$ represent its components in the ϕ coordinate basis, and ${}^\psi t^i$ its components in the ψ coordinate basis.

$$t = {}^\phi t^i \left. \frac{\partial_{--} \circ \phi^{\leftarrow}}{\partial x^i} \right|_{\phi(m)}$$

$$t = {}^\psi t^i \left. \frac{\partial_{--} \circ \psi^{\leftarrow}}{\partial x^i} \right|_{\psi(m)}$$

At the end of section 3, we obtained the tangent vector transformation law using chain rule. We did this by expressing one coordinate basis in terms of another.

$$\left.\frac{\partial_{--} \circ \phi^{\leftarrow}}{\partial x^i}\right|_{\phi(m)} = \left.\frac{\partial_{--} \circ \psi^{\leftarrow}}{\partial x^j}\right|_{\psi(m)} \left.\frac{\partial(\psi \circ \phi^{\leftarrow})^j}{\partial x^i}\right|_{\phi(m)}$$

$${}^{\psi}t^j = {}^{\phi}t^i \left.\frac{\partial(\psi \circ \phi^{\leftarrow})^j}{\partial x^i}\right|_{\phi(m)}$$

Now we're going to use the tangent vector transformation law to obtain the transformation law for the dual space. First we name the components of the cotangent vector $w \in \mathcal{T}^*_{m,M}$ in terms of the dual bases to the coordinate bases of ϕ and ψ.

$$w = {}^{\phi}w_i \; dx^i$$

$$w = {}^{\psi}w_i \; dy^i$$

We want to express the $\{dx^i\}$ basis in terms of $\{dy^i\}$. Look at the values of dx^i on the ψ coordinate basis:

$$\begin{aligned} dx^i \left(\left.\frac{\partial_{--} \circ \psi^{\leftarrow}}{\partial x^j}\right|_{\psi(m)} \right) &= dx^i \left(\left.\frac{\partial(\phi \circ \psi^{\leftarrow})^k}{\partial x^j}\right|_{\psi(m)} \left.\frac{\partial_{--} \circ \phi^{\leftarrow}}{\partial x^k}\right|_{\phi(m)} \right) \\ &= \left.\frac{\partial(\phi \circ \psi^{\leftarrow})^k}{\partial x^j}\right|_{\psi(m)} \left.\frac{\partial x^i \circ \phi^{\leftarrow}}{\partial x^k}\right|_{\phi(m)} \\ &= \left.\frac{\partial(\phi \circ \psi^{\leftarrow})^k}{\partial x^j}\right|_{\psi(m)} \delta^i_k \\ &= \left.\frac{\partial(\phi \circ \psi^{\leftarrow})^i}{\partial x^j}\right|_{\psi(m)} \end{aligned}$$

Then look at the values of $dy^k \left.\frac{\partial(\phi \circ \psi^{\leftarrow})^i}{\partial x^k}\right|_{\psi(m)}$ on the ψ coordinate basis:

$$\begin{aligned} \left[\left.\frac{\partial(\phi \circ \psi^{\leftarrow})^i}{\partial x^k}\right|_{\psi(m)} dy^k \right] \left(\left.\frac{\partial_{--} \circ \psi^{\leftarrow}}{\partial x^j}\right|_{\psi(m)} \right) &= \left.\frac{\partial(\phi \circ \psi^{\leftarrow})^i}{\partial x^k}\right|_{\psi(m)} \left.\frac{\partial y^k \circ \psi^{\leftarrow}}{\partial x^j}\right|_{\psi(m)} \\ &= \left.\frac{\partial(\phi \circ \psi^{\leftarrow})^i}{\partial x^k}\right|_{\psi(m)} \delta^k_j \\ &= \left.\frac{\partial(\phi \circ \psi^{\leftarrow})^i}{\partial x^j}\right|_{\psi(m)} \end{aligned}$$

They have the same values on the ψ coordinate basis, so they must be the same dual vector. Having expressed $\{dx^i\}$ in terms of $\{dy^i\}$, we may then relate the components of w in the different bases.

$$dx^i = dy^j \left.\frac{\partial(\phi \circ \psi^{\leftarrow})^i}{\partial x^j}\right|_{\psi(m)}$$

$${}^{\psi}w_i = {}^{\phi}w_j \left.\frac{\partial(\phi \circ \psi^{\leftarrow})^j}{\partial x^i}\right|_{\psi(m)}$$

How do we find the transformation law for an arbitrary tensor? The hardest part is actually over, this one is easy. I'll do an example with a $(1,2)$ type tensor, $A \in T^1_2(\mathcal{T}_{m,M})$. Remember from section 5.3 that the components of $A : V^* \times V \times V \to \mathbb{R}$ in a basis $\left\{ \frac{\partial_{--} \circ \psi^{\leftarrow}}{\partial x^i}\Big|_{\psi(m)} \otimes dy^j \otimes dy^k \right\}$ are

$$ {}^{\psi}A^i_{jk} = A\left(dy^i,\ \frac{\partial_{--} \circ \psi^{\leftarrow}}{\partial x^j}\Bigg|_{\psi(m)},\ \frac{\partial_{--} \circ \psi^{\leftarrow}}{\partial x^k}\Bigg|_{\psi(m)} \right) $$

Now we just use the vector and dual vector transformation laws that we got.

$$ = A\left(\frac{\partial(\psi \circ \phi^{\leftarrow})^i}{\partial x^a}\Bigg|_{\phi(m)} dx^a,\ \frac{\partial(\phi \circ \psi^{\leftarrow})^b}{\partial x^j}\Bigg|_{\psi(m)} \frac{\partial_{--} \circ \phi^{\leftarrow}}{\partial x^b}\Bigg|_{\phi(m)},\ \frac{\partial(\phi \circ \psi^{\leftarrow})^c}{\partial x^k}\Bigg|_{\psi(m)} \frac{\partial_{--} \circ \phi^{\leftarrow}}{\partial x^c}\Bigg|_{\phi(m)} \right) $$

Then we can just use multilinearity to take out the scalar factors, and we're left with the ϕ coordinate basis components of A:

$$ {}^{\psi}A^i_{jk} = {}^{\phi}A^a_{bc} \left(\frac{\partial(\psi \circ \phi^{\leftarrow})^i}{\partial x^a}\Bigg|_{\phi(m)} \frac{\partial(\phi \circ \psi^{\leftarrow})^b}{\partial x^j}\Bigg|_{\psi(m)} \frac{\partial(\phi \circ \psi^{\leftarrow})^c}{\partial x^k}\Bigg|_{\psi(m)} \right) $$

And that is our tensor transformation law. Applying this simple procedure to any tensor gives its tensor transformation law. Actually if one is not interested in rigorously setting up the geometry, one can deal only with the transformation laws. In fact, some people have no idea what manifolds or tangent spaces are, but they can still do highly geometrical physics because they understand transformation laws.

Chapter 6

Fields

This discussion follows the treatment of Bishop and Goldberg in chapter three, and it draws information from Schutz. If I were to provide a treatment of differential forms, this would be the time to do it. But I'm not going to treat them yet since they are a huge subject of their own and I'm currently in a poor position to be writing about them. This section is missing examples of solving for integral curves, so if you're not already familiar with them please try to do some problems yourself.

6.1 Vector Fields

A vector field on a chunk of a manifold should be an object that assigns to each point in the manifold a vector, an element of the tangent space there. A vector field X on an open $D \subset M$ is a function such that for any $m \in D$, we get some $X(m) \in \mathcal{T}_{m,M}$. Remember that a vector in a tangent space at m is actually a linear derivation-at-m, an operator defined by the way it acts on smooth functions around m. So a vector field returns a particular one of these operators at each point in $D \subset M$. This gives us another way to view (or define) a vector field. Consider a smooth function $f \in \mathcal{F}_{m,M}$ defined on some open $W \subset M$. Just as acting a vector on f gives a real number, we can consider acting a vector field on f to give a real number *at each point*. The vector field thus gives back some other function defined on $dom(f) \cap dom(X)$.

$$(\overline{X}(f))(m) = (X(m))(f) \quad \text{for all } m \in dom(f) \cap dom(X)$$

Although the different ways of looking at the vector field, X and $\overline{X}$, are strictly different objects, we glide between the different interpretations naturally enough to think of them as one object. I don't like this, I prefer that it's clear which function we're talking about. We might like our definition of vector fields to be analogous to our definition of vectors. In that case we could have directly defined them as operators on smooth real valued functions to real valued functions that are linear and obey a product rule.

We already defined the coordinate basis of a coordinate system ϕ at some point m in the coordinate system's domain U as the basis $\left\{ \left.\frac{\partial(\_\_\circ\phi^{\leftarrow})}{\partial x^i}\right|_{\phi(m)} \quad \text{for } i < \text{dimension} \right\}$. The mapping of $m \in U$ to the vector $\left.\frac{\partial(\_\_\circ\phi^{\leftarrow})}{\partial x^i}\right|_{\phi(m)}$ produces the i^{th} coordinate basis vector field ∂_i. To make our notation of ∂_i complete we would need to know what coordinate system we're talking about, so we may use ${}^{\phi}\partial_i$ if the need arises. Now our vector field X may be expressed in terms of the coordinate basis vector fields:

$$X = X^i \partial_i$$

where the X^i are functions on the manifold $X^i : dom(\phi) \cap dom(X) \to \mathbb{R}$, what we call the *component functions* of the vector field (note that the multiplication that appears above is the pointwise multiplication of a scalar function by a field of operators). This expansion into components works because at each $m \in dom(\phi) \cap dom(X)$, we may expand $X(m)$ in terms of the coordinate basis there. The actual components X^i are $\overline{X}(x^i)$ where x^i are the coordinate functions that make up ϕ:

$$\overline{X}(x^i) = \overline{(X^j \partial_j)}\ (x^i) = X^j\ \overline{\partial_j}(x^i) = X^i$$

What does it mean for a vector field to be smooth? A vector field is C^∞ (smooth) if for any smooth real valued function f on the manifold the function $\overline{X}(f)$ is also smooth. If a vector field is smooth then its components in a coordinate system must also be smooth. This is because the components are $\overline{X}(x^i)$ and the x^i are obviously smooth functions. Does the converse hold? Suppose the components of a vector field are smooth in every coordinate system on a manifold. Take any smooth $f : W \to \mathbb{R}$ on the manifold. The question to ask is whether $\overline{X}(f)$ is a smooth function for sure. Well at any point in $W \subset M$ there should be some coordinate system $\psi : U \to \mathbb{R}^n$ with $m \in U$. The function $\overline{X}(f)$ in this coordinate system has a domain $D \cap U \cap W$ and looks like $X^i\ \overline{{}^{\psi}\partial_i}(f)$. Since the X^i are smooth functions and the ${}^{\psi}\partial_i$ are smooth vector fields it is clear that $\overline{X}(f)$ is smooth. So *a vector field smooth iff it's components are smooth in all coordinate systems.*

6.2 Tensor Fields

A tensor field should assign a tensor to each point in some chunk of the manifold. A tensor field T of type (r, s) on some open subset D of the manifold is a function such that for every $m \in D$ we get some $T(m) \in T^r_s(\mathcal{T}_{m,M})$. Just like there were multiple ways to define a vector field, there is another kind of mapping that we can use to talk about a tensor field. But before we talk about that we need to understand dual vector fields. A vector field is clearly just a $(1, 0)$ type tensor field. A dual vector field is just a $(0, 1)$ type tensor field. A dual vector field $Y : D \to T^0_1(\mathcal{T}_{m,M})$ can be thought of as function that takes vector fields to real valued functions. That is, we may consider the alternative mapping $\overline{Y}$ which takes a vector field and gives a function defined by

$$(\overline{Y}(X))(m) = Y(m)(X(m)) \quad \text{for } m \in dom(X) \cap dom(Y)$$

Just like vector fields, these can be expanded in a basis at each point. The dual basis we've been talking about for $\mathcal{T}^*_{m,M}$ is denoted in a slightly misleading way, $\left\{dx^i\right\}$, where x^i are the coordinate functions of some coordinate system ϕ. A better notation, which I will now switch to using, is $\left\{dx^i_{(m)}\right\}$. This indicates that we're talking about the differential of the smooth function x^i *at a particular* $m \in M$. We're going to reserve the old notation for a field. We let $dx^i : dom(\phi) \to T^0_1(\mathcal{T}_{m,M})$ refer to the tensor *field* defined by $dx^i(m) = dx^i_{(m)}$. Now we can talk about the components of a dual vector field, which can be expanded in a coordinate basis field:

$$Y = Y_i dx^i$$

where the Y_i are real-valued functions on $dom(Y) \cap dom(\phi)$ and the sum and products seen above are of the pointwise function type. Actually the Y_i are just $\overline{Y}({}^{\phi}\partial_i)$, since

$$\overline{Y}({}^{\phi}\partial_i) = Y_j \overline{dx^j}({}^{\phi}\partial_i) = Y_i$$

Again, in practical situations nobody bothers with the difference between $\overline{Y}$ and Y; I'm just more comfortable pointing it out when we're doing this for the first time.

Tensors are multilinear functions on vectors and dual vectors, so a tensor field gives us a multilinear function on the tangent vectors and cotangent vectors at each point. Thus if we have a tensor field T, we can consider the function $\overline{T}$ that acts on vector fields and dual vector fields $(\omega^1, \ldots, \omega^r, \nu_1, \ldots, \nu_s)$ to produce real-valued functions defined by

$$\overline{T}(\omega^1, \ldots, \omega^r, \nu_1, \ldots, \nu_s)\ (m) = T(m)(\omega^1(m), \ldots, \omega^r(m), \nu_1(m), \ldots, \nu_s(m))$$

Tensor fields also have components in a basis. To express them this we way could define a "super tensor product" that acts on tensor fields. We can define the tensor field it makes out of tensor fields T and U by

$$(T \boxtimes U)(m) = T(m) \otimes U(m) \quad \text{for } m \in dom(T) \cap dom(U)$$

We may then express the tensor field T in terms of component functions, $T^{i_1 \ldots i_r}_{j_1 \ldots j_s}$, on a coordinate basis for ϕ with coordinate functions x^i:

$$T = T^{i_1 \ldots i_r}_{j_1 \ldots j_s}\ \partial_{i_1} \boxtimes \ldots \boxtimes \partial_{i_r} \boxtimes dx^{j_1} \boxtimes \ldots \boxtimes dx^{j_s}$$

The component functions can again be expressed as $T^{i_1 \ldots i_r}_{j_1 \ldots j_s} = \overline{T}(dx^{i_1}, \ldots, dx^{i_r}, \partial_{j_1}, \ldots, \partial_{j_s})$

What does it mean for a tensor field to be smooth? We already said that a vector field is smooth if feeding it any smooth function produces a smooth function. We then showed that this was equivalent to having all the components of a vector field be smooth in every coordinate basis. Let's see if we can say similar things about tensor fields. A $(0,1)$ type tensor field, a dual vector field, is smooth (C^∞) if feeding it any smooth vector field produces a smooth function. An (r,s) type tensor field is smooth if feeding it r smooth dual vector fields and s smooth vector fields produces a smooth function. We now obtain similar statements about the smoothness of component functions. We will summarize them below. Consider a vector field X, a dual vector field Y, and an (r,s) type tensor field T:

$X \in C^\infty \implies X^i \in C^\infty$ in any coordinate system because $X^i = \overline{X}(x^i)$ and $x^i \in C^\infty$

$X^i \in C^\infty$ in any coordinate system $\implies X \in C^\infty$ because $X = X^i \partial_i$ and $\partial_i \in C^\infty$

$Y \in C^\infty \implies Y_i \in C^\infty$ in any coordinate system because $Y_i = \overline{Y}(\partial_i)$ and $\partial_i \in C^\infty$

$Y_i \in C^\infty$ in any coordinate system $\implies Y \in C^\infty$ because $Y = Y_i dx^i$ and $dx^i \in C^\infty$

$T \in C^\infty \implies T^{i_1 \ldots i_r}_{j_1 \ldots j_s} \in C^\infty$ in any coordinate system because $T^{i_1 \ldots i_r}_{j_1 \ldots j_s} = \overline{T}(dx^{i_1}, \ldots, dx^{i_r}, \partial_{j_1}, \ldots, \partial_{j_s})$ and $dx^{i_1}, \ldots, dx^{i_r}, \partial_{j_1}, \ldots, \partial_{j_s} \in C^\infty$

$T^{i_1 \ldots i_r}_{j_1 \ldots j_s} \in C^\infty$ in any coordinate system $\implies T \in C^\infty$ because $T = T^{i_1 \ldots i_r}_{j_1 \ldots j_s}\ \partial_{i_1} \boxtimes \ldots \boxtimes \partial_{i_r} \boxtimes dx^{j_1} \boxtimes \ldots \boxtimes dx^{j_s}$ and $\partial_{i_1}, \ldots, \partial_{i_r}, dx^{j_1}, \ldots, dx^{j_s} \in C^\infty$

6.3 Integral Curves

An integral curve $\gamma \colon \mathbb{R} \to M$ of a vector field X is a path in the manifold whose tangent everywhere is the value of the field. That is, γ is an integral curve of X if $(\_ \circ \gamma)' = X \circ \gamma$ (hidden in this statement is the requirement that $ran(\gamma) \subset dom(X)$). We say that the integral curve "starts at" m if $\gamma(0) = m$. It should be clear that the definition of integral curve has nothing to say about starting point. So what happens when we reparameterize the curve? Take an integral curve γ defined on the real interval $]\!]a, b[\![$. Consider a reparametrization of γ's variable, $r :]\!]c, d[\![\to]\!]a, b[\![$. Then $(\_\_ \circ \gamma \circ r)' = (\_\_ \circ \gamma) \circ r \cdot r'$. If we had $r' = 1$, a constant 1 function, then we'd have $(\_\_ \circ \gamma \circ r)' = (\_\_ \circ \gamma)' \circ r = X \circ \gamma \circ r$. The fact that $(\_\_ \circ (\gamma \circ r))' = X \circ (\gamma \circ r)$ tells us that the reparameterized curve, $\gamma \circ r$, is also an integral curve. When is $r' = 1$? This tells us that we can translate the parameter of an integral curve and we'll still have an integral curve. It also tells us that *specifying the value of an integral curve at one point completely determines it.* (Well, almost. Restricting the domain of an integral curve specified in this way does also make an integral curve. But we often just want the "longest" curve we can get.)

Suppose that someone hands you a vector field and asks you for the integral curve that starts at a particular point in your manifold. To solve this problem, you reach into your atlas and grab an appropriate coordinate system ϕ with coordinate functions x^i. Let us express the requirement that $(\_\_ \circ \gamma)' = X \circ \gamma$ in terms of ϕ. For $s \in dom(\gamma)$:

$$(\_\_ \circ \gamma)'(s) = (\_\_ \circ \phi^{\leftarrow} \circ \phi \circ \gamma)'(s) = \left.\frac{\partial \_\_ \circ \phi^{\leftarrow}}{\partial x^j}\right|_{\phi(\gamma(s))} \cdot (\phi \circ \gamma)^{j\,\prime}(s) = (\partial_j \circ \gamma)(s) \cdot (x^j \circ \gamma)'(s)$$

$$(\_\_ \circ \gamma)' = (x^j \circ \gamma)' \cdot (\partial_j \circ \gamma)$$
$$X \circ \gamma = (X^j \partial_j) \circ \gamma = (X^j \circ \gamma) \cdot (\partial_j \circ \gamma)$$

The requirement to be an integral curve is then $(x^j \circ \gamma)' \cdot (\partial_j \circ \gamma) = (X^j \circ \gamma) \cdot (\partial_j \circ \gamma)$. For every $s \in dom(\gamma)$ and $\gamma(s) \in dom(\phi)$, $(\partial_j \circ \gamma)(s)$ is a basis vector. So by requiring equality of vector components, the condition becomes a system of differential equations:

$$(x^j \circ \gamma)' = X^j \circ \gamma$$

This says that the derivative of the j^{th} component of γ as viewed through ϕ is equal to the j^{th} component of the vector field at each point in the path. It is a system of first order ordinary differential equations. The starting point gives us the initial conditions so that it has a unique solution for $x^j \circ \gamma$ (again, unique up to a restriction of the domain). It could be that $dom(\phi)$ does not cover $dom(X)$. In that case we'd solve the differential equations for several coordinate systems and glue the pieces together.

I'm now going to skip a lot of differential equation theory. I've already skipped some by stating without proof that the system of equations above has a unique solution. I'm going to skip some powerful theorems about how big the domains and ranges of our integral curves could be, the reason being that they are too mathematical for what I want to get to right now. The bottom line is that any integral curve for a smooth vector field defined on all of a compact manifold can have its domain extended to all of $\mathbb{R}$. Any vector field with the property that all its integral curves can be extended to $\mathbb{R}$ is called *complete*.

6.4 Flows

Flows add no new information to what integral curves give us; they are just a different way of looking at the same thing. If you're standing in a vector field, you can travel along an integral curve by just walking along the arrows you're standing on with the appropriate speed. So we have one point in the manifold (say m), some "time" t passes, and then we're at another point in the manifold (say p). A particular integral curve is generated once we pick an m, it maps each t to a p (the integral curve starting at m). Similarly, part of a flow is generated when we pick a particular t, we get something that maps each m to p. The whole flow is the object that produces a map from the m's to the p's for each choice of t. So the flow is one object that summarizes the information contained in all of the integral curves of a vector field. Let's make this a little more formal.

The flow of a vector field X is a function $\widetilde{X}$ that maps real numbers to functions from the manifold to itself in the following way. Given an $s \in \mathbb{R}$, $\widetilde{X}_s \vdots M \to M$ such that for any $m \in dom(\widetilde{X}_s)$, $\widetilde{X}_s(m)$ is $\gamma_m(s)$ where γ_m is the integral curve of X that starts at m. So $\widetilde{X}_s$ is a function that pushes each point on the manifold along an integral curve by parameter s. Back in the day, people used to talk about vector fields as generators of infinitesimal transformations. People thought of pushing points on the manifold just a little bit along all the arrows. We see that "infinitesimal" values of s make $\widetilde{X}_s$ seem like an "infinitesimal" transformation. However, we are above using such language; $\widetilde{X}$ is the vector field's flow.

The existence and uniqueness theorems that we skipped for integral curves follow us here. It is their results that ensure that if X is complete then $dom(\widetilde{X}_s) = dom(X)$ for $s \in \mathbb{R}$. Notice that $\widetilde{X}_0$ is the identity map on the manifold. And $\widetilde{X}_s \circ \widetilde{X}_t$ does intuitively seem like $\widetilde{X}_{s+t}$. Properties like these make $\widetilde{X}$ a *one parameter group*.

A one parameter group is not a group! It is a continuous homomorphism from $\mathbb{R}$ to a (topological) group. Our continuous homomorphism is the flow itself, $\widetilde{X}$. And the group is its range, $ran(\widetilde{X})$. The group operation is composition of the functions. We have an identity ($\widetilde{X}_0$), an inverse ($\widetilde{X}_s \mapsto \widetilde{X}_{-s}$), associativity, and commutativity. That $\widetilde{X}$ is homomorphic comes from $\widetilde{X}_s \circ t\widetilde{X}_t = \widetilde{X}_{s+t}$. To see proofs of this, continuity, and smoothness I refer the reader to Bishop and Goldberg page 125 or Warner page 37. Again there is too much differential equation theory that I don't want to go into. It's enough to just know that the flows of smooth vector fields are smooth in their variables (that is, the mapping $s, m \mapsto \widetilde{X}_s(m)$ is smooth).

Chapter 7

Lie Derivatives

This section attempts to combine the intuitive motivation of Schutz, the computational practicality of Bishop and Goldberg, and the mathematical rigor of Warner. Some of the terminology is my own. I'm going to do most of the discussion on Lie derivatives of vector fields, to avoid cluttering the calculations. Forms and other tensor fields come from vectors anyway, and it will not be too hard to generalize at the end. The usual notation for evaluating a function a point, "$f(x)$," fails us here; there is just way too much going on. To make calculations easier on the eyes I'll be using the alternative notation, "$\langle f \mid x\rangle$".

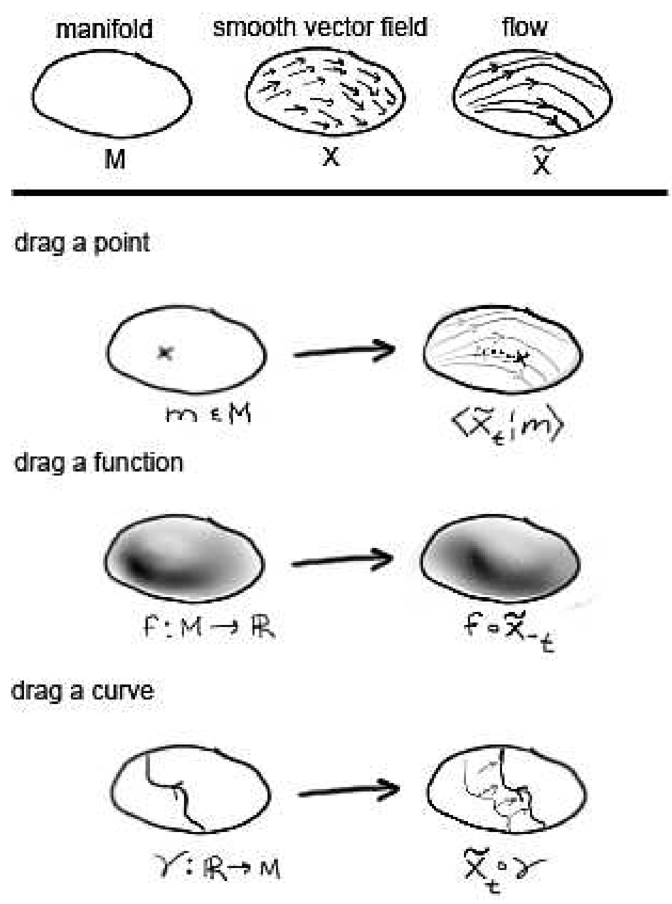
7.1 Dragging

Good terminology can go a long way when it comes to visualizing advanced concepts. Here we introduce the notion of Lie dragging. In section 6.4 we examined the one parameter group $\widetilde{X}$ made from the C^∞ vector field X. Since each $\widetilde{X}_t$ has an inverse, $\widetilde{X}_{-t}$, and both are smooth functions from the manifold to itself, we know that each $\widetilde{X}_t$ is a diffeomorphism. We may use it to take one point in the manifold, m, to another point along an X integral curve, $\langle \widetilde{X}_t \mid m\rangle$. We will call this procedure "Lie dragging m along the vector field X." The amount of dragging being done is given by the parameter t.

We can drag a real-valued function $f \vdots M \to \mathbb{R}$ in the following way. At each point we want the value of the dragged function to be the value of the function pushed forward along the integral curves of X, so we look backwards by parameter amount t and return the function's value there. The Lie dragged function is then $f \circ \widetilde{X}_{-t}$.

We can drag a curve $\gamma : \mathbb{R} \to M$ by simply dragging the points in the range of the curve. The Lie dragged curve along X is then $\widetilde{X}_t \circ \gamma$.

Now this one is the whopper. How can we Lie drag a vector field? We need a way to take all the vectors from some vector field Y along the integral curves of the C^∞ vector field X. That is, we need a way of knowing what a vector looks like in a foreign tangent space, given that we used the diffeomorphism $\widetilde{X}_t$ to take it there. Remember that cold definition in section 4.1? This is a job for the differential map; it's time to understand it.



When we defined it in section 4.1, we only needed to use the differential in the special case where the second manifold (the range of the diffeomorphism) is $\mathbb{R}$. We were identifying $\mathcal{T}_{m,\mathbb{R}}$ with $\mathbb{R}$ for each $m \in \mathbb{R}$, so we weren't really noticing the tangent space to tangent space action of the differential.

When not dealt with carefully, pictures often serve to destroy our mathematical intuition rather than reinforce it. Upon every utterance of the words "vector field," the image of scattered arrows is conjured in my head. What even gives us the right to draw an arrow for a vector?! The tangent vector is nothing more than an operator on real-valued functions smoothly defined around some point. In pictorial representations all the information contained in that operator is reduced to a dinky little arrow. What we are really drawing is the direction in which real-valued functions should change to get a big real number when we feed them to the tangent vector; and the scale of the arrow just scales that real number. The picture is an abstraction of the mathematical object. It's an abstraction of an abstraction.

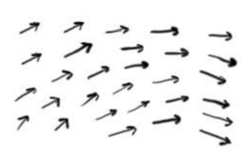


Figure 7.1: Is this a vector field? No. It's a picture.

Keeping in mind that that little arrow is just an action on functions, we reapproach the definition of the differential. Consider the diffeomorphism $\psi : M \to M$. $d\psi$ should take us from $\mathcal{T}_{m,M}$ to $\mathcal{T}_{\langle \psi \mid m \rangle, M}$ using ψ. For some $v \in \mathcal{T}_{m,M}$, $\langle d\psi \mid v \rangle$ is just an action on functions. It makes perfect sense to take whatever function is fed to it, drag it backwards using ψ so the smooth part is back at m, and then feed that to v. This intuitively takes the action on functions (vector) in one place to the same action on functions (vector) dragged forward using ψ, by simply dragging back all the functions using ψ! Indeed we defined $\langle \langle d\psi \mid v \rangle \mid f \rangle = \langle v \mid f \circ \psi \rangle$ for $f \in \mathcal{F}_{\langle \psi \mid m \rangle, M}$ (remember that dragging the function forward would have used $\psi^{\leftarrow}$).

So to Lie drag a vector $v \in \mathcal{T}_{m,M}$ along the vector field X by amount t we just feed it to the differential: $\langle d\widetilde{X}_t \mid v \rangle$. We can Lie drag a vector field Y along the vector field X in the following way. At each point we wish to look back by integral curve parameter amount t and return the vector there. However we need that vector to be in the right tangent space, otherwise the result would not really be a vector field (it would be some horrible function that returns a foreign tangent vector at each point). So we drag it along by t using the differential. The Lie dragged vector field is then $d\widetilde{X}_t \circ Y \circ \widetilde{X}_{-t}$.

(Note: We noted before that the notation $d\psi$ was incomplete, that $d\psi_{m,M}$ would be more appropriate. I'm just using $d\psi$ as if it worked for all points, like a mapping between tangent bundles instead of just between tangent spaces. Set-theoretically, I'm taking $d\psi$ to be the union of all the $d\psi_{m,M}$ for $m \in dom(\psi)$)

In case you are still not convinced that this was the "right way" to define the dragging of a vector field, I have one more trick up my sleeve. Let us consider what happens to the integral curves of a vector field when we Lie drag it along another vector field. Consider smooth vector fields X and Y (we will drag Y along X by amount t). Let γ represent arbitrary integral curves of Y. Let γ^* represent arbitrary integral curves of the dragged field, $d\widetilde{X}_t \circ Y \circ \widetilde{X}_{-t}$. We are searching for a relationship between the γ's and the γ^*'s.

Each γ is a solution to:

$$Y \circ \gamma = (\_\_ \circ \gamma)'$$

And each γ^* is a solution to:

$$\begin{aligned}
& & d\widetilde{X}_t \circ Y \circ \widetilde{X}_{-t} \circ \gamma^* &= (__ \circ \gamma^*)' \\
& & \langle d\widetilde{X}_t \circ Y \circ \widetilde{X}_{-t} \circ \gamma^* \mid s \rangle &= \langle (__ \circ \gamma^*)' \mid s \rangle \\
\forall\, f \in \mathcal{F}_{\langle \gamma^* \mid s \rangle, M} & & \langle \langle d\widetilde{X}_t \mid \langle Y \mid \langle \widetilde{X}_{-t} \circ \gamma^* \mid s \rangle \rangle \rangle \mid f \rangle &= \langle \langle (__ \circ \gamma^*)' \mid s \rangle \mid f \rangle \\
\forall\, f \in \mathcal{F}_{\langle \gamma^* \mid s \rangle, M} & & \langle \langle Y \mid \langle \widetilde{X}_{-t} \circ \gamma^* \mid s \rangle \rangle \mid f \circ \widetilde{X}_t \rangle &= \langle (f \circ \gamma^*)' \mid s \rangle \\
\forall\, f \in \mathcal{F}_{\langle \gamma^* \mid s \rangle, M} & & \langle \langle Y \mid \langle \widetilde{X}_{-t} \circ \gamma^* \mid s \rangle \rangle \mid f \circ \widetilde{X}_t \rangle &= \langle (f \circ \widetilde{X}_t \circ \widetilde{X}_{-t} \circ \gamma^*)' \mid s \rangle \\
\forall\, g \in \mathcal{F}_{\langle \widetilde{X}_{-t} \mid \langle \gamma^* \mid s \rangle \rangle, M} & & \langle \langle Y \mid \langle \widetilde{X}_{-t} \circ \gamma^* \mid s \rangle \rangle \mid g \rangle &= \langle (g \circ \widetilde{X}_{-t} \circ \gamma^*)' \mid s \rangle \\
\forall\, g \in \mathcal{F}_{\langle \widetilde{X}_{-t} \mid \langle \gamma^* \mid s \rangle \rangle, M} & & \langle \langle Y \mid \langle \widetilde{X}_{-t} \circ \gamma^* \mid s \rangle \rangle \mid g \rangle &= \langle \langle (__ \circ \widetilde{X}_{-t} \circ \gamma^*)' \mid s \rangle \mid g \rangle \\
& & \langle Y \circ \widetilde{X}_{-t} \circ \gamma^* \mid s \rangle &= \langle (__ \circ \widetilde{X}_{-t} \circ \gamma^*)' \mid s \rangle \\
& & Y \circ (\widetilde{X}_{-t} \circ \gamma^*) &= (__ \circ (\widetilde{X}_{-t} \circ \gamma^*))'
\end{aligned}$$

So $\widetilde{X}_{-t} \circ \gamma^* = \gamma$, an integral curve of Y. Or rather $\gamma^* = \widetilde{X}_t \circ \gamma$. The integral curves of the dragged vector field are just the draggings of the integral curves of the original vector field! In fact we could have just as well *defined* the Lie dragging of Y along X in the following way: Take the integral curves of Y, drag them all along X by amount t, and return the vector field associated with the new set of integral curves. This result is a nice and simple statement to fall back on if you one day completely forget how to visualize Lie derivatives (see figure 7.2).

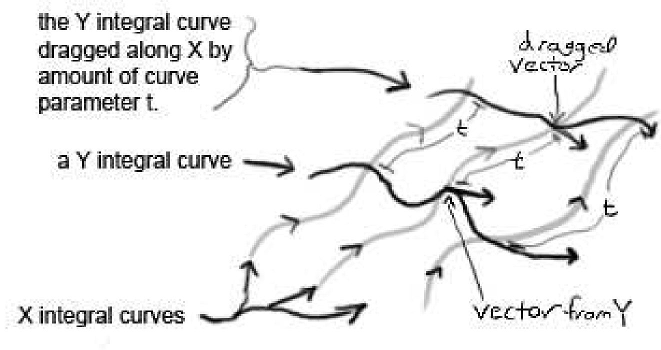


Figure 7.2: Read the labels starting from the bottom.

One last bit of terminology: *Backdragging* something by t means dragging it by $-t$. And when a field doesn't change upon Lie dragging, it is said to be *drag-invariant*.

7.2 The Idea

In section 6.1 we saw how a C^∞ vector field X can be viewed as a mapping from C^∞ functions to C^∞ functions, $\overline{X}$. Just like a tangent vector is a linear derivation-at-a-point, $\overline{X}$ is a linear derivation. It takes the directional derivative at each point of the function that is fed to it, and gives back the field of directional derivatives. $\langle \overline{X} \mid f \rangle$ is the "derivative of f along X." The Lie derivative is an attempt to generalize the C^∞ function to C^∞ function derivative-like action of $\overline{X}$ to a C^∞ vector field to C^∞ vector field derivative-like action (and eventually for any rank tensor field).

Suppose I have a field of stuff defined on a manifold, $☺ : M \to$ stuff, and I want to quantify how it changes along my smooth vector field X. We begin by picking a point $m \in M$, and asking how the field of stuff, ☺, changes along X *at that point*. The natural thing to do would be to step forward from m a little in the direction dictated by X (to $\langle \widetilde{X}_t \mid m \rangle$), look at the field of stuff there, $\langle ☺ \mid \langle \widetilde{X}_t \mid m \rangle \rangle$, and subtract from that the value of the field at m, $\langle ☺ \mid \langle \widetilde{X}_t \mid m \rangle \rangle - \langle ☺ \mid m \rangle$. But we know from experience that it's best not to move away from m during this procedure; field values at m don't always play nicely with field values somewhere else (like if ☺ was a vector field). So instead of stepping forward a little, we backdrag the entire field and evaluate that at m. We then make it a genuine derivative by dividing by the size of the "step" and taking a limit:

$$\langle £_X ☺ \mid m \rangle = \lim_{\text{drag amount} \to 0} \frac{\langle ☺ \text{ backdragged} \mid m \rangle - \langle ☺ \mid m \rangle}{\text{amount of dragging}}$$

And we call that the Lie derivative of ☺ with respect to X at m. Now let's do it for real...

7.3 The Definition

The Lie derivative of a smooth vector field Y with respect to a smooth vector field X is the vector field $\pounds_X Y$ defined as follows.

$$\langle \pounds_X Y \mid m \rangle = \lim_{t \to 0} \frac{\langle d\widetilde{X}_{-t} \circ Y \circ \widetilde{X}_t \mid m \rangle - \langle Y \mid m \rangle}{t}$$

It is the limit of a $\mathcal{T}_{m,M}$-valued function at each point m. According to the previous section, all we need to do to figure out how to define the Lie derivative of something is figure out how to drag it. For example, let's apply the same concept to a real-valued function on the manifold, $f : M \to \mathbb{R}$. The previous section tells us to define it's Lie derivative as:

$$\langle \pounds_X f \mid m \rangle = \lim_{t \to 0} \frac{\langle f \circ \widetilde{X}_t \mid m \rangle - \langle f \mid m \rangle}{t}$$

Let γ_m be the integral curve of X that starts at m. Then

$$\begin{aligned}
\langle \pounds_X f \mid m \rangle &= \lim_{t \to 0} \frac{\langle f \circ \gamma_m \mid t \rangle - \langle f \circ \gamma_m \mid 0 \rangle}{t} \\
&= \langle (f \circ \gamma_m)' \mid 0 \rangle \\
&= \langle \langle (__ \circ \gamma_m)' \mid 0 \rangle \mid f \rangle \\
&= \langle \langle X \circ \gamma_m \mid 0 \rangle \mid f \rangle \\
&= \langle \langle X \mid m \rangle \mid f \rangle \\
&= \langle \langle \overline{X} \mid f \rangle \mid m \rangle
\end{aligned}$$

So $\pounds_X f = \langle \overline{X} \mid f \rangle$. This is a pleasant result; it reinforces the notion that the Lie derivative $\pounds_X$ generalizes the action of $\overline{X}$ on real-valued functions.

7.4 Other Ways of Looking at It

$\pounds_X Y$ has a pretty simple appearance in a drag-invariant basis. Suppose $\{\hat{E}_i\}$ is a drag-invariant basis field. That is, suppose that for each i we had $\hat{E}_i = d\widetilde{X}_t \circ \hat{E}_i \circ \widetilde{X}_{-t}$ for all t. (One could obtain such a basis field by starting out with a set of basis vectors at some point on each integral curve of X, then dragging them around to define the rest of the basis field. So if I know $\langle \hat{E}_i \mid m \rangle$ then I may define $\langle \hat{E}_i \mid \langle \widetilde{X}_t \mid m \rangle \rangle = \langle d\widetilde{X}_t \mid \langle \hat{E}_i \mid m \rangle \rangle$ and this would force $\hat{E}_i$ to be drag invariant). Now express Y in this basis: $Y = Y^i \hat{E}_i$, where $Y^i \vdots M \to \mathbb{R}$ are component functions. In the following lines we use the linearity of the differential (which is easy to prove straight from the definition in section 4.1). We also use the fact that $(d\psi)^{\leftarrow} = d(\psi^{\leftarrow})$, which is also straight-forward to prove.

$$
\begin{aligned}
\langle \pounds_X Y \mid m\rangle &= \lim_{t\to 0} \frac{\langle d\widetilde{X}_{-t} \mid \langle Y \mid \langle \widetilde{X}_t \mid m\rangle\rangle\rangle - \langle Y \mid m\rangle}{t} \\
&= \lim_{t\to 0} \frac{\langle d\widetilde{X}_{-t} \mid \langle Y^i \mid \langle \widetilde{X}_t \mid m\rangle\rangle \cdot \langle \hat{E}_i \mid \langle \widetilde{X}_t \mid m\rangle\rangle\rangle - \langle Y \mid m\rangle}{t} \\
&= \lim_{t\to 0} \frac{\langle Y^i \mid \langle \widetilde{X}_t \mid m\rangle\rangle \cdot \langle d\widetilde{X}_{-t} \mid \langle \hat{E}_i \mid \langle \widetilde{X}_t \mid m\rangle\rangle\rangle - \langle Y \mid m\rangle}{t} \\
&= \lim_{t\to 0} \frac{\langle Y^i \mid \langle \widetilde{X}_t \mid m\rangle\rangle \cdot \langle d\widetilde{X}_{-t} \mid \langle d\widetilde{X}_t \mid \langle \hat{E}_i \mid m\rangle\rangle\rangle - \langle Y^i \mid m\rangle \cdot \langle \hat{E}_i \mid m\rangle}{t} \\
&= \lim_{t\to 0} \frac{\langle Y^i \mid \langle \widetilde{X}_t \mid m\rangle\rangle - \langle Y^i \mid m\rangle}{t} \cdot \langle \hat{E}_i \mid m\rangle \\
&= \langle \hat{E}_i \mid m\rangle \lim_{t\to 0} \frac{\langle Y^i \mid \langle \widetilde{X}_t \mid m\rangle\rangle - \langle Y^i \mid m\rangle}{t}
\end{aligned}
$$

The limit that appears here is just, by definition, $\langle \pounds_X Y^i \mid m\rangle$. According to what we previously showed, that's just $\langle\langle \overline{X} \mid Y^i\rangle \mid m\rangle$. So the Lie derivative takes the following simple form in our drag-invariant basis field:

$$\pounds_X Y = \langle \overline{X} \mid Y^i\rangle \cdot \hat{E}_i$$

This is yet another way to have defined Lie derivatives. We could have said that the Lie derivative of a vector field with respect to X is the derivative of its components along X in a drag-invariant basis field. When we extend our definition this statement will also hold for the components of a tensor field in a drag-invariant basis field. Another interesting calculation is to examine $\pounds_{\partial_1} Y$, the Lie derivative with respect to the first coordinate basis field for some coordinate system ϕ.

Consider the basis field $\{\partial_i\}$ for a coordinate system ϕ with coordinate functions x^i. First I claim that ∂_i is a drag-invariant basis field with respect to ∂_1. We showed that dragging a vector field is the same thing as taking its integral curves, dragging them, and looking at the vector field associated with the new curves. This means the drag-invariance of a vector field is equivalent to the drag invariance of the *set* of integral curves. The integral curves of ∂_i are easy to solve for in the ϕ coordinate system. The result is $\langle \phi \circ \widetilde{(\partial_i)}_t \mid m\rangle = (\langle x^1 \mid m\rangle, \langle x^2 \mid m\rangle, \ldots, \langle x^i \mid m\rangle + t, \ldots, \langle x^n \mid m\rangle)$. (Just write down the differential equations for the components and the initial condition then solve, as explained in section 6.3). So the flow of ∂_i just adds t to the i^{th} coordinate of a point. In particular the flow of ∂_1 just adds t to the first coordinate of a point. Lie dragging a whole integral curve of ∂_i, $\gamma \mapsto \widetilde{(\partial_1)}_t \circ \gamma$, would then just add t (a constant) to the first coordinate of the curve, which obviously produces some other integral curve of ∂_i. So Lie dragging takes integral curves of ∂_i to integral curves of ∂_i, leaving the whole set of integral curves drag-invariant. According to what we said this means that ∂_i is drag-invariant with respect to ∂_1. Using the results from the previous calculation, we may now write the Lie derivative in a simple way:

$$\pounds_{\partial_1} Y = \langle \overline{\partial_1} \mid Y^i\rangle \cdot \partial_i$$

where $\langle \overline{\partial_1} \mid Y^i\rangle$ is of course just the real-valued function $\frac{\partial Y^i \circ \phi^{\leftarrow}}{\partial x^1}$. We have shown that $\pounds_{\partial_1}$ takes the derivative of all the components of a vector field with respect to the first coordinate! This property of $\pounds_{\partial_1}$ will also be valid for tensor components when we extend our definition. Let's do that now.

7.5 Extending the Definition

Section 7.2 has already laid out the conceptual framework behind defining Lie derivatives. All that remains is to figure out how to drag any rank tensor field along a vector field. Let's start with dual vector fields. Recall that we were able to drag a vector (an action on functions) along a diffeomorphism by backdragging the functions it acts on. We essentially moved the vector through a diffeomorphism by moving its effect on functions. Our dragging tool was thus the differential map, though we don't really call it dragging unless the diffeomorphism comes from the flow of a vector field.

Since a dual vector w is just an action on vectors, we may move the dual vector through a diffeomorphism by moving its effect on vectors, which we now know how to move with the differential. So the dragged dual vector is that map which upon being fed a vector backdrags it and feeds it to the original dual vector. For a diffeomorphism $\psi : M \to M$ and some $m \in M$ we define $D\psi : \mathcal{T}^*_{m,M} \to \mathcal{T}^*_{\langle \psi \,\vdots\, m\rangle,M}$ such that for any dual vector $w \in \mathcal{T}^*_{m,M}$,

$$\langle\langle D\psi \,\vdots\, w\rangle \,\vdots\, v\rangle = \langle w \,\vdots\, \langle d\psi^{\leftarrow} \,\vdots\, v\rangle\rangle \qquad \text{for all } v \in \mathcal{T}_{\langle \psi \,\vdots\, m\rangle,M}$$

(Note: Again $D\psi_{m,M}$ or $D\psi_m$ would be better notation, but I'll mush together all the $D\psi_m$'s when I use $D\psi$. Also be warned that authors like to define the inverse of this map instead of what I've done, and they call it $\delta\psi$. I don't see the point of this, we would just have to keep taking inverses later on. So that's why I'm using capital D instead of δ.)

Now we have the ability to Lie drag dual vectors along vector fields, $w \mapsto \langle D\widetilde{X}_t \,\vdots\, w\rangle$. And similarly we may Lie drag dual vector fields, $W \mapsto D\widetilde{X}_t \circ W \circ \widetilde{X}_{-t}$. We then define the Lie derivative of a dual vector field according to the same scheme:

$$\langle \pounds_X W \,\vdots\, m\rangle = \lim_{t\to 0} \frac{\langle D\widetilde{X}_{-t} \circ W \circ \widetilde{X}_t \,\vdots\, m\rangle - \langle W \,\vdots\, m\rangle}{t}$$

You've probably already figured out how we're going to Lie drag a tensor along a vector field. Being a multilinear operator on vectors and dual vectors, a tensor may be moved through a diffeomorphism by using the diffeomorphism to move its effect on vectors and dual vectors, both of which we now know how to move. For a diffeomorphism $\psi : M \to M$ and some $m \in M$ we define $D^r_s\psi : T^r_s(\mathcal{T}_{m,M}) \to T^r_s(\mathcal{T}_{\langle \psi \,\vdots\, m\rangle,M})$ such that for any (r,s) type tensor, $T \in T^r_s(\mathcal{T}_{m,M})$ we have:

$$\langle\langle D^r_s\psi \,\vdots\, T\rangle \,\vdots\, (w_1,\ldots,w_r,v_1,\ldots,v_s)\rangle = \langle T \,\vdots\, \ (\langle D\psi^{\leftarrow} \,\vdots\, w_1\rangle,\ldots,\langle D\psi^{\leftarrow} \,\vdots\, w_r\rangle,\langle d\psi^{\leftarrow} \,\vdots\, v_1\rangle,\ldots,\langle d\psi^{\leftarrow} \,\vdots\, v_s\rangle)\ \rangle$$

for all $w_1,\ldots,w_r \in \mathcal{T}^*_{\langle \psi \,\vdots\, m\rangle,M}$ and $v_1,\ldots,v_s \in \mathcal{T}_{\langle \psi \,\vdots\, m\rangle,M}$

We may now Lie drag (r,s) type tensors, $T \mapsto \langle D^r_s\widetilde{X}_t \,\vdots\, T\rangle$.

And we may Lie drag (r,s) type tensor fields, $U \mapsto D^r_s\widetilde{X}_t \circ U \circ \widetilde{X}_{-t}$.

And we may define the Lie derivative of an (r,s) tensor field U with respect to a smooth vector field X:

$$\langle \pounds_X U \,\vdots\, m\rangle = \lim_{t\to 0} \frac{\langle D^r_s\widetilde{X}_{-t} \circ U \circ \widetilde{X}_t \,\vdots\, m\rangle - \langle U \,\vdots\, m\rangle}{t}$$

So $\pounds_X U$ is another (r,s) tensor field defined in terms of the limit of a tensor-valued function at each point. Now you are in a good position to go back and verify the properties we claimed would still hold after extending the definition. For example, $\pounds_X U$ takes the form $\langle \overline{X} \mathrel{\vdots} U^{i_1 \ldots i_r}_{j_1 \ldots j_s} \rangle \cdot \hat{E}_{i_1} \boxtimes \ldots \boxtimes \hat{E}_{i_r} \boxtimes \hat{\Sigma}^{j_1} \boxtimes \ldots \boxtimes \hat{\Sigma}^{j_s}$ if U is expanded in a drag-invariant basis field $\{\hat{E}_i\}$ with a corresponding dual basis field $\{\hat{\Sigma}^i\}$. Just use what you learned in section 6.2 and follow the calculation in 7.4 to prove this.

7.6 Properties of Lie Derivatives

There are a couple of expected properties whose proofs are straight-forward. The first is distributivity of $\pounds_X$ over addition.

$$\langle \pounds_X \mathrel{\vdots} A + B \rangle = \langle \pounds_X \mathrel{\vdots} A \rangle + \langle \pounds_X \mathrel{\vdots} B \rangle$$

for tensor fields A and B. The proof is a trivial consequence of the definition. The other property is Leibniz rule (product rule behavior).

$$\langle \pounds_X \mathrel{\vdots} A \boxtimes B \rangle = \langle \pounds_X \mathrel{\vdots} A \rangle \boxtimes B + A \boxtimes \langle \pounds_X \mathrel{\vdots} B \rangle$$

I found a couple of ways to prove this. One is to take the limit definition at a point, and feed the tensor arbitrary vectors and dual vectors. This turns tensor products into multiplication. The proof can be finished in the same way that product rule is proven for regular $\mathbb{R} \to \mathbb{R}$ functions (add and subtract an appropriate term). Another way to prove it is to use the extended version of the result obtained in 7.4. Getting a drag-invariant basis and working with the components makes it pretty trivial to prove.

The property we're about to prove is pretty exciting. The *Lie bracket* of two smooth vector fields X and Y is defined as follows.

$$[X, Y] = \overline{X} \circ \overline{Y} + \overline{Y} \circ \overline{X}$$

should be −

The usual next step would be to show that $[X, Y]$ is in fact a vector field (which should be a little surprising since is a second order operator). We would do this by proving that it's a linear derivation; some nice canceling removes the second order effects. But this follows from what we're about to prove anyway, so I don't need to show it. I unfortunately did not have time to get into the details of Lie brackets, but I refer the reader to Schutz pages 43-49. The Lie bracket has a very important geometrical interpretation that Schutz does a great job explaining.

7.6.1 Proof That $\overline{\pounds_X Y} = [X, Y]$

Consider two smooth vector fields X and Y. To prove this, we need to show that $\langle \overline{\pounds_X Y} \mathrel{\vdots} f \rangle = \langle [X, Y] \mathrel{\vdots} f \rangle$ for smooth real-valued functions f defined on $dom(x) \cap dom(Y)$. Refining it even more, we need to show that

$$\langle \langle \pounds_X Y \mathrel{\vdots} m \rangle \mathrel{\vdots} f \rangle = \langle \langle \overline{X} \mathrel{\vdots} \langle \overline{Y} \mathrel{\vdots} f \rangle \rangle \mathrel{\vdots} m \rangle - \langle \langle \overline{Y} \mathrel{\vdots} \langle \overline{X} \mathrel{\vdots} f \rangle \rangle \mathrel{\vdots} m \rangle$$

In this proof we will repeatedly use the fact that

$$\begin{aligned} \langle \langle \overline{X} \mathrel{\vdots} f \rangle \mathrel{\vdots} m \rangle &= \langle \pounds_X f \mathrel{\vdots} m \rangle \\ &= \lim_{t \to 0} \frac{\langle f \circ \widetilde{X}_t \mathrel{\vdots} m \rangle - \langle f \mathrel{\vdots} m \rangle}{t} = \left. \frac{d}{dt} \right|_{t=0} \langle f \circ \widetilde{X}_t \mathrel{\vdots} m \rangle \end{aligned}$$

for any smooth vector field and scalar field X and f. We already showed this in section 7.3. We'll start by rewriting what we need to show, then we'll work on the left side of the equal sign.

$$\begin{aligned}
\langle\langle \overline{X} \mid \langle \overline{Y} \mid f\rangle\rangle \mid m\rangle - \langle\langle \overline{Y} \mid \langle \overline{X} \mid f\rangle\rangle \mid m\rangle &= \left.\frac{d}{dt}\right|_{t=0} \langle\langle \overline{Y} \mid f\rangle \mid \langle \widetilde{X}_t \mid m\rangle\rangle - \left.\frac{d}{du}\right|_{u=0} \langle\langle \overline{X} \mid f\rangle \mid \langle \widetilde{Y}_u \mid m\rangle\rangle \\
&= \left.\frac{d}{dt}\right|_{t=0} \left.\frac{d}{du}\right|_{u=0} \langle f \circ \widetilde{Y}_u \circ \widetilde{X}_t \mid m\rangle - \left.\frac{d}{du}\right|_{u=0} \left.\frac{d}{ds}\right|_{s=0} \langle f \circ \widetilde{X}_s \circ \widetilde{Y}_u \mid m\rangle \\
&= \left.\frac{\partial^2}{\partial t \partial u}\right|_{(t,u)=(0,0)} \langle f \circ \widetilde{Y}_u \circ \widetilde{X}_t \mid m\rangle - \left.\frac{\partial^2}{\partial s \partial u}\right|_{(u,s)=(0,0)} \langle f \circ \widetilde{X}_s \circ \widetilde{Y}_u \mid m\rangle
\end{aligned}$$

We used the smoothness of our maps to switch the order of the partial derivatives in the last step. Of course the choices of letters for parameters like $t,\ u,\ s$ were arbitrary. The notation "$\frac{d}{dt}$" is informal anyway. I just chose letters that will match the result we get at the end. What we must show has been reduced to:

$$\langle\langle \pounds_X Y \mid m\rangle \mid f\rangle = \left.\frac{\partial^2}{\partial t \partial u}\right|_{(t,u)=(0,0)} \langle f \circ \widetilde{Y}_u \circ \widetilde{X}_t \mid m\rangle - \left.\frac{\partial^2}{\partial s \partial u}\right|_{(u,s)=(0,0)} \langle f \circ \widetilde{X}_s \circ \widetilde{Y}_u \mid m\rangle$$

Let's get to work on the left hand side.

$$\begin{aligned}
\langle\langle \pounds_X Y \mid m\rangle \mid f\rangle &= \left\langle \lim_{t\to 0} \frac{\langle d\widetilde{X}_{-t} \circ Y \circ \widetilde{X}_t \mid m\rangle - \langle Y \mid m\rangle}{t} \mid f \right\rangle \\
&= \lim_{t\to 0} \frac{\langle\langle d\widetilde{X}_{-t} \circ Y \circ \widetilde{X}_t \mid m\rangle \mid f\rangle - \langle\langle Y \mid m\rangle \mid f\rangle}{t} \\
&= \lim_{t\to 0} \frac{\langle\langle Y \mid \langle \widetilde{X}_t \mid m\rangle\rangle \mid f \circ \widetilde{X}_{-t}\rangle - \langle\langle Y \mid m\rangle \mid f\rangle}{t} \\
&= \lim_{t\to 0} \frac{\langle\langle \overline{Y} \mid f \circ \widetilde{X}_{-t}\rangle \mid \langle \widetilde{X}_t \mid m\rangle\rangle - \langle\langle \overline{Y} \mid f\rangle \mid m\rangle}{t} \\
&= \left.\frac{d}{dt}\right|_{t=0} \langle\langle \overline{Y} \mid f \circ \widetilde{X}_{-t}\rangle \mid \langle \widetilde{X}_t \mid m\rangle\rangle \\
&= \left.\frac{d}{dt}\right|_{t=0} \left.\frac{d}{du}\right|_{u=0} \langle f \circ \widetilde{X}_{-t} \circ \widetilde{Y}_u \mid \langle \widetilde{X}_t \mid m\rangle\rangle \\
&= \left.\frac{\partial^2}{\partial t \partial u}\right|_{(t,u)=(0,0)} \langle f \circ \widetilde{X}_{-t} \circ \widetilde{Y}_u \circ \widetilde{X}_t \mid m\rangle
\end{aligned}$$

Here's the tricky bit. We're going to define a couple of functions so we can formalize away from the "$\frac{d}{dt}$" notation. This will help us to use the chain rule correctly. Define a function $\kappa \vdots \mathbb{R}^3 \to \mathbb{R}$ as follows.

$$\langle \kappa \mid (t,u,s)\rangle = \langle f \circ \widetilde{X}_s \circ \widetilde{Y}_u \circ \widetilde{X}_t \mid m\rangle$$

So the thing being differentiated above is the mapping $(t,u) \mapsto \langle \kappa \,\text{¦}\, (t,u,-t)\rangle$. Notice that κ is well-defined on some neighborhood of $(0,0,0)$. Define $j : \mathbb{R}^2 \to \mathbb{R}^3$ to be the mapping $(t,u) \mapsto (t,u,-t)$ with component functions $\langle j^1 \,\text{¦}\, (t,u)\rangle = t$, $\langle j^2 \,\text{¦}\, (t,u)\rangle = u$, and $\langle j^3 \,\text{¦}\, (t,u)\rangle = -t$. Now the thing being differentiated above is $\kappa \circ j$. Let number subscripts denote partial derivatives. Then we have

$$\langle\langle \pounds_X Y \,\text{¦}\, m\rangle \,\text{¦}\, f\rangle = \langle (\kappa \circ j)_{21} \,\text{¦}\, (0,0)\rangle$$

Using chain rule (ⓜ is matrix multiplication):

$$\begin{aligned}
(\kappa \circ j)' &= \begin{bmatrix} (\kappa \circ j)_1 & (\kappa \circ j)_2 \end{bmatrix} \\
&= \begin{bmatrix} \kappa_1 \circ j & \kappa_2 \circ j & \kappa_3 \circ j \end{bmatrix} ⓜ \begin{bmatrix} j^1_1 & j^2_1 \\ j^1_2 & j^2_2 \\ j^1_3 & j^2_3 \end{bmatrix} \\
&= \begin{bmatrix} \kappa_1 \circ j & \kappa_2 \circ j & \kappa_3 \circ j \end{bmatrix} ⓜ \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & 0 \end{bmatrix}
\end{aligned}$$

So the first derivative gives

$$(\kappa \circ j)_2 = \begin{bmatrix} \kappa_1 \circ j & \kappa_2 \circ j & \kappa_3 \circ j \end{bmatrix} ⓜ \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \kappa_2 \circ j$$

And similarly the next derivative gives

$$(\kappa \circ j)_{21} = (\kappa_2 \circ j)_1 = \begin{bmatrix} \kappa_{21} \circ j & \kappa_{22} \circ j & \kappa_{23} \circ j \end{bmatrix} ⓜ \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} = \kappa_{21} \circ j - \kappa_{23} \circ j$$

And now we may switch back to the less formal notation:

$$\begin{aligned}
\langle\langle \pounds_X Y \,\text{¦}\, m\rangle \,\text{¦}\, f\rangle &= \langle \kappa_{21} \circ j - \kappa_{23} \circ j \,\text{¦}\, (0,0)\rangle \\
&= \langle \kappa_{21} - \kappa_{23} \,\text{¦}\, (0,0,0)\rangle \\
&= \frac{\partial^2}{\partial t \partial u}\bigg|_{(t,u,s)=(0,0,0)} \langle f \circ \widetilde{X}_s \circ \widetilde{Y}_u \circ \widetilde{X}_t \,\text{¦}\, m\rangle - \frac{\partial^2}{\partial s \partial u}\bigg|_{(t,u,s)=(0,0,0)} \langle f \circ \widetilde{X}_s \circ \widetilde{Y}_u \circ \widetilde{X}_t \,\text{¦}\, m\rangle \\
&= \frac{\partial^2}{\partial t \partial u}\bigg|_{(t,u)=(0,0)} \langle f \circ \widetilde{Y}_u \circ \widetilde{X}_t \,\text{¦}\, m\rangle - \frac{\partial^2}{\partial s \partial u}\bigg|_{(u,s)=(0,0)} \langle f \circ \widetilde{X}_s \circ \widetilde{Y}_u \,\text{¦}\, m\rangle \\
&= \langle\langle [X,Y] \,\text{¦}\, f\rangle \,\text{¦}\, m\rangle
\end{aligned}$$

Thus $\overline{\pounds_X Y} = [X,Y]$ and the proof is done.

Chapter 8

Appendix

8.1 Notation

I describe the notation I use within the text where I first use it. However I'm providing a list here in case you want to skip around the paper.

- $\mathcal{F}_{m,M}$ is the set of real-valued functions smoothly defined on a neighborhood of a point m on the manifold M.
- $\mathcal{T}_{m,M}$ is the tangent space at the point m in the manifold M.
- V^* is the dual space to the vector space V.
- $T^r_s V$ is the vector space of (r, s) type tensors over the vector space V.
- $dom(f)$ and $ran(f)$ are the domain and range of a function f.
- ${}^\phi\partial_i$ is the i^{th} coordinate basis *field* that comes from the coordinate system ϕ. Sometimes I leave out the ϕ.
- $f \vdots A \to B$ is just like $f : A \to B$ except the domain of f is some subset of A.
- Be careful of the differential (like dx^i or $d\widetilde{X}_t$), they tend to have meanings that change in subtle ways. Their particular usage is described in the text each time.
- $\overline{X}$ is the C^∞ function to function way of looking at a vector field X. Most people just use X to refer to both.
- $\widetilde{X}$ is the flow of the smooth vector field X. When I say flow I mean the one-parameter group associated with the vector field.
- $\langle f \mid x \rangle$ is an alternative (and better) notation for $f(x)$, evaluating a function at a point. Since it's unconventional I avoided it as much as I could, but I was forced to use it at the end.
- The underscore notation is sometimes used to denote a function. For example the parabola defined by $f(x) = x^2$ is just $f = \_\_^2$. And the tangent vector defined by $t(f) = (f \circ \gamma)'$ is $t = (\_\_ \circ \gamma)'$