MATH 622: DIFFERENTIAL GEOMETRY I

DEAN BASKIN

1. Introduction

This is the first semester of a two-semester graduate course providing an introduction to differential geometry.

We are mostly using Lee's Smooth Manifolds book but will supplement with others.

Topics roughly covered include:

- Examples of manifolds and submanifolds
- Differentiable manifolds and smooth maps between them
- Tangent and cotangent vectors and bundles, differentials, vector fields, Lie brackets
- Tensors
- Distributions and the Frobenius theorem
- Differential forms and integration; de Rham cohomology
- Lie groups
- Classical differential geometry of curves and surfaces
- Gauss–Bonnet

There are many paths to differential geometry. Common ones include algebraic geometry, algebraic topology, and general relativity. Another common path involves pizza (this is a joke about Gauss's Theorema Egregium).

2. Preliminaries

2.1. **Topology and topological manifolds.** The main objects of study in this course are smooth manifolds. Before we define them, we'll first define topological manifolds. A naïve first definition is that a topological manifold is a topological space that "locally looks like $\mathbb{R}^n$".

Definition 1. A topological space is a set X equipped with a collection $\mathcal{T} \subset \mathcal{P}(X)$ so that

(1) $\mathcal{T}$ contains the empty set and the whole space: $\emptyset, X \in \mathcal{T}$;
(2) $\mathcal{T}$ is closed under finite intersections: if $U_1, \dots, U_n \in \mathcal{T}$, then $U_1 \cap \dots \cap U_n \in \mathcal{T}$.
(3) $\mathcal{T}$ is closed under arbitrary unions: if $\{U_i\}_{i\in I} \subset \mathcal{T}$, then $\cup_{i\in I} U_i \in \mathcal{T}$.

Elements of $\mathcal{T}$ are called open sets. A set $F \subset X$ is closed if $U = X \setminus F$ is open.

Suppose X and Y are topological spaces.

Definition 2. A function $f : X \to Y$ is continuous if for every open set $V \subset Y$, the preimage $f^{-1}(V)$ is open.

Definition 3. A continuous function $f : X \to Y$ is a homeomorphism if it is a bijection and its inverse is continuous.

Definition 4. A topological space X is locally Euclidean if for every point $p \in X$ there is an open subset $U \subset X$ with $p \in U$ so that U is homeomorphic to an open subset of $\mathbb{R}^n$ for some n.

First working definition of a topological manifold: A topological manifold is a locally Euclidean topological space.

This definition seems to capture what we mean by "locally looks like $\mathbb{R}^n$", so why isn't this our definition? Well, bad things can happen here. Here are two examples that can be fleshed out.

- The line with two origins.
- The very long line.
- An uncountable set with the discrete topology.

One way to avoid these pitfalls is to add some conditions. Spivak uses metrizability, while Lee uses the following definition:

Definition 5. A manifold is a topological space M that is

(1) Hausdorff, i.e., for any $p, q \in M$, there are disjoint open neighborhoods of p and q;
(2) Second-countable, i.e., there is a countable base for its topology; and
(3) Locally Euclidean.

We won't prove the following theorem. The fastest proofs I know of go through homology.

Theorem 6 (Invariance of domain)**.** *If $U \subset \mathbb{R}^n$ is open and $f : U \to \mathbb{R}^n$ is continuous and injective then $f(U)$ is open.*

One implication of this theorem is that the n in the definition is determined uniquely by the point and so is determined on each connected component of M. Throughout our course, we'll assume it's the same on each connected component and say $n = \dim M$.

Aside: connectivity.

Definition 7. A topological space X is connected if, given any open sets A, B with $A, B \neq \emptyset$, $A \cup B = X$, we must have $A \cap B \neq \emptyset$.

In other words, X is connected if the only sets that are both open and closed are $\emptyset$ and X. This splits X into connected components.

Many other things can go wrong. We'll ask that our manifolds be σ-compact (i.e., a countable union of compact sets).

2.2. **Calculus.** We now review some facts from multivariable calculus.

Definition 8. Suppose $U \subset \mathbb{R}^n$ is open. A function $f : U \to \mathbb{R}^m$ is differentiable at a point $a \in U$ if there is a linear map $L_a : \mathbb{R}^n \to \mathbb{R}^m$ so that

$$\lim_{x \to a} \frac{|f(x) - f(a) - L_a(x-a)|}{|x-a|} = 0.$$

We say that the linear map L_a is the (total) derivative of f at a.

If f is differentiable at all points in U, one can view the total derivative of f as a function $U \to \mathrm{Lin}(\mathbb{R}^n, \mathbb{R}^m)$, with the latter space being isomorphic to $\mathbb{R}^{n \times m}$. This total derivative is denoted by a variety of different notations, including f' and Df.

One immediate consequence of the definition of the derivative is the following version of "Taylor's theorem":

Theorem 9. *Suppose $U \subset \mathbb{R}^n$ is open, $f : U \to \mathbb{R}^m$ is differentiable at $a \in U$ and W is any convex subset of U, then*

$$f(x) = f(a) + Df(a)(x-a) + R(x),$$

where

$$\lim_{x \to a} \frac{|R(x)|}{|x-a|} = 0$$

The chain rule then reads: If $f : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at p and $g : \mathbb{R}^m \to \mathbb{R}^k$ is differentiable at $f(p)$, then $g \circ f : \mathbb{R}^n \to \mathbb{R}^k$ is differentiable at p with derivative $Dg \circ Df$ (with composition in the sense of linear maps).

If $f : U \to \mathbb{R}^m$ is continuous on U, differentiable on U, and its total derivative is a continuous function, we say that $f \in C^1(U; \mathbb{R}^m)$ or that f is C^1. When the codomain is $\mathbb{R}$ we typically drop the codomain from the notation. Defining higher derivatives via the chain rule is then straightforward; if f and its first k derivatives are all continuous on U then we say f is in C^k. If f and all of its derivatives are continuous, we say f is C^∞; if in addition f is an analytic function, we sometimes say f is C^ω. General functions in this course will be C^∞ unless otherwise specified.

The j-th partial derivative of a function $\phi : \mathbb{R}^n \to \mathbb{R}$ is given by

$$\partial_j \phi(x) = \frac{\partial \phi}{\partial x^j}(x) = \lim_{h \to 0} \frac{\phi(x + he_j) - \phi(x)}{h},$$

where e_j is the j-th standard basis vector in $\mathbb{R}^n$.

Picking bases for $\mathbb{R}^n$ and $\mathbb{R}^m$ provides a matrix representation of the total derivative of a map $f : \mathbb{R}^n \to \mathbb{R}^m$. Indeed, a basis for $\mathbb{R}^m$ lets us write

$$f = \begin{pmatrix} f^1 \\ f^2 \\ \dots \\ f^m \end{pmatrix},$$

and then the total derivative Df has matrix

$$\begin{pmatrix} \partial_1 f^1 & \partial_2 f^1 & \dots & \partial_n f^1 \\ \partial_1 f^2 & \partial_2 f^2 & \dots & \partial_n f^2 \\ \vdots & \vdots & \ddots & \vdots \\ \partial_1 f^m & \partial_2 f^m & \dots & \partial_n f^m \end{pmatrix}.$$

Theorem 10 (Taylor's theorem)**.** *Let $U \subset \mathbb{R}^n$ be open and $a \in U$. Let $f \in C^{k+1}(U)$ for some $k \geq 0$. If V is any convex subset of U containing a then for all $x \in V$,*

$$f(x) = P_k(x) + R_k(x),$$

where P_k is the k-th order Taylor polynomial of f at a, i.e.,

$$P_k(x) = \sum_{m=0}^{k} \frac{1}{m!} \sum_{|\alpha|=m} \partial_\alpha f(a)(x-a)^\alpha,$$

and R_k is the remainder term given by

$$R_k(x) = \frac{1}{k!} \sum_{|\alpha|=k+1} (x-a)^\alpha \int_0^1 (1-t)^k (\partial_\alpha f)(a + t(x-a))\, dt.$$

In particular, if all of the $(k+1)$-st partial derivatives of f are bounded by M then for all $x \in W$,

$$|f(x) - P_k(x)| \leq \frac{n^{k+1}M}{(k+1)!}|x-a|^{k+1}$$

Proof. Let's do it first for the one-dimensional version. Suppose $U \subset \mathbb{R}$, $g \in C^{k+1}(U)$, $W \subset U$ is an interval and $a \in W$. The Taylor polynomial of g is given by

$$P_k(x) = \sum_{m=0}^{k} \frac{1}{m!} g^{(m)}(a)(x-a)^m,$$

while the remainder term is given by

$$R_k(x) = \frac{1}{k!}(x-a)^{k+1} \int_0^1 (1-t)^k g^{(k+1)}(a + t(x-a))\, dt.$$

We'll prove it by induction. The statement for $k = 0$ follows immediately from the fundamental theorem of calculus and the chain rule applied to the function $g'(a + t(x-a))$. Now we assume it holds for some k and integrate the remainder $R_k(x)$ by parts:

$$\begin{aligned} R_k(x) &= \frac{1}{k!}(x-a)^{k+1}\left(-\frac{(1-t)^{k+1}}{k+1} g^{(k+1)}(a + t(x-a))|_0^1 + \frac{x-a}{k+1}\int_0^1 (1-t)^{k+1} g^{(k+2)}(a+t(x-a))\, dt \right) \\ &= \frac{1}{(k+1)!}(x-a)^{k+1} g^{(k+1)}(a) + R_{k+1}(x), \end{aligned}$$

so that indeed $f(x) = P_{k+1}(x) + R_{k+1}(x)$ for all $x \in W$. The final statement in one dimension follows by bounding the integrand by $M(1-t)^k$ and integrating.

For the more general statement we use the mutlivariable chain rule and the one-variable statement. (This is a homework assignment.) □

One way to think about Taylor's theorem for C^∞ functions is as a kind of divisibility theorem: If $f \in C^\infty$ vanishes as a point a, then f is in the C^∞-span of $(x^1 - a^1), (x^2 - a^2), \ldots, (x^n - a^n)$. In other words, the C^∞-ideal of functions vanishing at the point a is generated by these n functions.

Two of the most important results from multivariable calculus (though you probably did not learn them in a multivariable calculus class) are the inverse and implicit function theorems. We'll use the inverse function theorem repeatedly when talking about charts and the implicit function theorem will provide us with an essentially endless source of examples of smooth manifolds.

Theorem 11 (Inverse function theorem)**.** *Suppose $U \subset \mathbb{R}^n$ is open, $f \in C^1(U; \mathbb{R}^n)$, and $A = (Df)_p$ is the total derivative of f at p for some $p \in U$. If A is invertible (viewed as a linear map), then there is an open ball $B(p,r) \subset U$ with $r > 0$ so that*

(a) f is one-to-one on $B = B(p,r)$,
(b) $V = f(B)$ is open, and
(c) $g = (f|_B)^{-1}$ is C^1 on V.
(d) In addition, if f is C^k, so is g.

Proof. We start by writing

$$f(x) = f(p) + A(x-p) + R(x).$$

As f is C^1, we know that $R(x)$ is also C^1. Moreover, as we know that

$$\lim_{x\to p} \frac{|R(x)|}{|x-p|} = 0,$$

given any $\epsilon > 0$, we may find $r > 0$ so that

$$|R(x)| \leq \epsilon |x - p|, \quad |DR(x)| \leq \epsilon$$

for all $x \in B(p, r)$. In particular, for all $x, y \in B(p, r)$, we have

$$\begin{aligned} |R(x) - R(y)| &= \left| \int_0^1 (DR)(y + t(x-y))(x-y)\, dt \right| \\ &\leq \epsilon |x - y|. \end{aligned}$$

We now claim that by choosing $\epsilon > 0$ we guarantee that f is one-to-one on $B = B(p, r)$. Because A is invertible, there is some $c > 0$ (one can take c to be the size of the eigenvalue of A closest to 0) so that for all $x, y \in \mathbb{R}^n$,

$$|A(x-y)| \geq c|x-y|.$$

We now take $\epsilon = c/2$ and observe that for $x, y \in B(p, r)$,

$$\begin{aligned} |f(x) - f(y)| &= |A(x-y) + R(x) - R(y)| \\ &\geq |A(x-y)| - |R(x) - R(y)| \geq c|x-y| - \frac{c}{2}|x-y| \geq \frac{c}{2}|x-y|, \end{aligned}$$

so that f is injective on $B = B(p, r)$. Moreover, as invertibility is an open condition and f is C^1, we can ensure (by possibly shrinking r) that $f'(x)$ is invertible on all $x \in B$.

Now, letting $V = f(B)$, we want to show that V is open. Pick $y_0 \in V$, so there is some $x_0 \in B$ with $f(x_0) = y_0$. Choose $t > 0$ so that $\overline{B(x_0, t)} \subset B$, and take $\ell(x) = |f(x) - y_0|$ as a function on the sphere $|x - x_0| = t$. As f is one-to-one here and the sphere is compact, ℓ attains a positive minimum m here. Take $|y_1 - y_0| < m/3$, so that

$$|f(x) - y_1| \geq |f(x) - y_0| - |y_0 - y_1| \geq \frac{2m}{3}.$$

Now let $h(x) = |f(x) - y_1|^2$ on $\overline{B(x_0, t)}$. The previous calculation implies that $h(x) \geq 4m^2/9$ on the boundary of $\overline{B(x_0, t)}$, while $h(x_0) = |y_0 - y_1|^2 < m^2/9$, so the minimum of h is attained on the interior of $B(x_0, t)$. Let $x_1 \in B(x_0, t)$ denote the point where this minimum is attained. As x_1 is therefore a critical point of h, we have

$$0 = \nabla h(x_1) = 2f'(x_1)^{\intercal}(f(x_1) - y_1).$$

We know that $f'(x_1)$ is invertible, so $f(x_1) = y_1$ and thus $B(y_0, m/3) \subset V$ and thus V is open.

We now show that $g = (f|_B)^{-1}$ is C^1 on V. (Its derivative should be $f'(g(y))^{-1}$.) Fix $y \in V$ and for $|k|$ small with $y + k \in V$, we find $x, x + h \in B$ so that $y = f(x)$ and $y + k = f(x + h)$. Setting $B = (f'(g(y))^{-1}$, we have

$$\frac{|g(y+k) - g(y) - Bk|}{|k|} = \frac{|x + h - x - B(f(x+h) - f(x))|}{|k|} = \frac{|-B\left(f(x+h) - f(x) - f'(x)h\right)|}{|k|}.$$

If we show that $|k| \geq c\,|h|$ for some $c > 0$ independent of h, then this quotient will go to 0 as $k \to 0$. Indeed, by using Taylor's theorem around p again we see that

$$\begin{aligned}|k| &= |f(x+h) - f(x)| = |Ah + R(x+h) - R(x)| \\ &\geq |Ah| - |R(x+h) - R(x)| \geq c\,|h| - \frac{c}{2}\,|x+h-x| = \frac{c}{2}\,|h|\,.\end{aligned}$$

We may therefore take a limit as $h \to 0$ to see that g is differentiable at y with derivative $f'(g(y))^{-1}$. As this function is continuous, the derivative of g is also continuous.

Finally, if f is C^k, then the chain rule applied to g shows that g is also C^k. □

In discussing (and proving) the implicit function theorem it is helpful to think about splitting coordinates on the product space $\mathbb{R}^{n+m}$. For $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$, we'll use (x, y) for a point in $\mathbb{R}^{n+m}$. Similarly, if $A : \mathbb{R}^{n+m} \to \mathbb{R}^m$ is linear, we'll split it into (A_x, A_y), where $A_x : \mathbb{R}^n \to \mathbb{R}^m$ and $A_y : \mathbb{R}^m \to \mathbb{R}^m$ are given by

$$A_x(v) = A(v, 0), \quad A_y(w) = A(0, w).$$

Theorem 12 (Implicit function theorem)**.** *Suppose $f : E \to \mathbb{R}^m$ is C^1, where $E \subset \mathbb{R}^n \times \mathbb{R}^m$ is open. Suppose further that $f(p, q) = 0$ for some $(p, q) \in E$. Let $A = f'(p, q)$. If $A_y = (\partial_{y_j} f)_{j=1}^m$ is invertible, then there are open sets $U \subset \mathbb{R}^{n+m}$ and $W \subset \mathbb{R}^n$ so that*

(a) for each $x \in W$ there is a unique y so that $(x, y) \in U$ and $f(x, y) = 0$, and

(b) letting $g(x)$ denote this y, then $g : W \to \mathbb{R}^m$ is C^1, $g(p) = q$, $f(x, g(x)) = 0$, and $g'(p) = -(A_y)^{-1}A_x$.

To get the formula for the derivative right, it is instructive to think about the (trivial) linear version of the theorem:

Theorem 13 (Implicit function theorem, linear version)**.** *If $A : \mathbb{R}^{n+m} \to \mathbb{R}^m$ is linear and A_y is invertible, then for every $v \in \mathbb{R}^n$, there is a unique $w \in \mathbb{R}^m$ so that $A(v, w) = 0$.*

Proof. This is just the observation that

$$A(v, w) = A_x v + A_y w,$$

so that if this is to equal zero we must have

$$w = -A_y^{-1} A_x v.$$

□

Proof of implicit function theorem. For $(x, y) \in E$, we define a new function $F : E \to \mathbb{R}^{n+m}$ by $F(x, y) = \begin{pmatrix} x \\ f(x, y) \end{pmatrix}$. Note that F is C^1 and DF has the following matrix at (p, q):

$$F'(p, q) = \begin{pmatrix} I_{n\times n} & 0 \\ \partial_{x_j} f(p, q) & \partial_{y_\ell} f(p, q) \end{pmatrix} = \begin{pmatrix} I_{n\times n} & 0 \\ A_x & A_y \end{pmatrix}.$$

As A_y is invertible, $F'(p, q)$ is also invertible, so by the inverse function we may find an open cube $U \subset \mathbb{R}^{n+m}$ so that F is one-to-one on U, $V = F(U)$ is open, and $(F|_U)^{-1}$ is C^1.

We now let $W = \{x \in \mathbb{R}^n \mid (x, 0) \in V\}$. Note that $p \in W$ because $f(p, q) = 0$ and that W is open because V is open. If $x \in W$, then $(x, 0) \in V$ and so there is some y with $(x, y) \in U$ with $f(x, y) = 0$.

This y is our candidate for $g(x)$; we must show it is unique. Indeed, if $(x, y') \in U$ also satisfies $f(x, y') = 0$, then

$$F(x, y') = (x, f(x, y')) = (x, f(x, y)) = F(x, y),$$

so $y' = y$ as F is injective on U. We now define $y = g(x)$ for $x \in W$ to be the unique y with $(x, y) \in U$ and $f(x, y) = 0$, i.e., so that $F(x, g(x)) = (x, 0)$ for all $y \in W$.

Letting $G = (F|_U)^{-1}$, we know that G is C^1 and $(x, g(x)) = G(x, 0)$, so g is also C^1. We finally calculate the derivative of g. Indeed, let $\Phi : \mathbb{R}^n \to \mathbb{R}^{n+m}$ be given by $\Phi(x) = (x, g(x))$, so then as a linear map, we have

$$\Phi'(x)v = (v, g'(x)v) \quad \text{for all } v \in \mathbb{R}^n.$$

Now, since $f(\Phi(x)) = 0$, we must have $f'(\Phi(x)) \circ \Phi'(x) = 0$, i.e., at (p, q) we have

$$0 = f'(p, q)(I, g'(p)) = A_x + A_y g'(p),$$

and so $g'(p) = -A_y^{-1} A_x$. □

The implicit function theorem will provide us with many examples of manifolds more or less "for free"; rather than taking the trouble of defining charts (to be discussed below), we instead can check the rank of a map to verify that the level set must be a manifold. Indeed, it can be viewed as giving you local homeomorphisms (as we'll see, you get something even better) putting a locally Euclidean structure on the level set.

Corollary 14. *If $f : E \to \mathbb{R}^m$ is C^∞, where $E \subset \mathbb{R}^n \times \mathbb{R}^m$ is open, $f(p, q) = 0$, and $A = f'(p, q)$ has rank m, then the level set $f^{-1}(0)$ is locally Euclidean in a neighborhood of (p, q).*

As a quick example, consider the sphere $\mathbb{S}^n = \{\sum_{j=1}^{n+1} (x^j)^2 = 1\} \subset \mathbb{R}^{n+1}$. This set is the zero set of the function $f : \mathbb{R}^{n+1} \to \mathbb{R}$ given by

$$f(x^1, x^2, \ldots, x^{n+1}) = \sum_{j=1}^{n+1} (x^j)^2 - 1.$$

Its total derivative at a point on the sphere is given by

$$Df_{(x^1, x^2, \ldots, x^{n+1})} = \begin{pmatrix} 2x^1 & 2x^2 & \ldots & 2x^{n+1} \end{pmatrix},$$

which has rank one as long as at least one of the x^j is non-vanishing (as it must be to lie on the sphere). The sphere is therefore a topological manifold.

3. Smooth manifolds and maps between them

3.1. **Definitions.** Returning back to the main thread, we'd like to endow topological manifolds with smooth structures, i.e, with a rule for determining which functions are differentiable and how to differentiate them. Here's a naïve idea for what we'd like to do. We know that for every point in M there is a neighborhood U and a homeomorphism $\phi : U \to \phi(U) \subset \mathbb{R}^n$. We'd like to declare that $f : M \to \mathbb{R}$ is differentiable if $f \circ \phi^{-1} : \phi(U) \to \mathbb{R}$ is differentiable for every such pair (ϕ, U). This is roughly the right idea, but there's a significant problem: You typically can't differentiate homeomorphisms. (Indeed, if $f : \mathbb{R}^n \to \mathbb{R}^n$ is a homeomorphism it is "typically" differentiable *nowhere.*) If $\psi : V \to \psi(V)$ is another such homeomorphism and $U \cap V \neq \emptyset$,

$$f \circ \psi^{-1} = (f \circ \phi^{-1}) \circ (\phi \circ \psi^{-1}) : \mathbb{R}^n \to \mathbb{R}$$

need not be differentiable.

Here's our fix: we'll restrict our homeomorphisms to the class for which $\phi \circ \psi^{-1}$ is always differentiable.

We're typically interested in the C^∞ category, and I'll use "smooth" to mean this. Many of the results, on the other hand, apply in weaker or stronger settings and I encourage you to think about when they fail.

Notational conventions: now we'll typically use letters like x and y for these local homeomorphisms and we'll call them "coordinates". If U is an open subset of M for which $x : U \to x(U) \subset \mathbb{R}^n$ is a homeomorphism like this, we'll typically refer to (x, U) as a "chart". (The field is full of mapmaking analogies.)

Definition 15. If U, V are open subsets of M equipped with homeoemorphisms

$$\begin{aligned} x :& U \to x(U) \subset \mathbb{R}^n, \\ y :& V \to y(V) \subset \mathbb{R}^n, \end{aligned}$$

we say that (x, U) and (y, U) are C^∞-related if the "transition functions" are smooth, i.e., if

$$\begin{aligned} x \circ y^{-1} &: y(U \cap V) \to x(U \cap V), \\ y \circ x^{-1} &: x(U \cap V) \to y(U \cap V) \end{aligned}$$

are both smooth.

Definition 16. A family of C^∞-related homeomorphisms whose domains cover M is an *atlas* for M. A member (x, U) of an atlas is called a chart or a coordinate system for U.

Intuitive picture: it provides a way of assigning coordinates $(x^1(p), x^2(p), \ldots, x^n(p))$ to the points $p \in U$ so that you can treat it like $\mathbb{R}^n$.

If there is another homeomorphism that is C^∞-compatible with all of the charts in an atlas $\mathcal{A}$, you can create a new larger atlas $\mathcal{A}'$ by including it. Perhaps unsurprisingly, Zorn's lemma gives the following result:

Lemma 17. *If $\mathcal{A}$ is an atlas on M, $\mathcal{A}$ is contained in a unique maximal atlas $\mathcal{A}'$ for M.*

We can then define a smooth (or C^k, etc) manifold:

Definition 18. A C^∞ manifold (also called a smooth manifold) is a pair $(M, \mathcal{A})$, where M is a topological manifold and $\mathcal{A}$ is a maximal atlas for M.

Since each atlas is contained in a unique maximal atlas, specifying the manifold requires only specifying some atlas.

Examples

(1) $(\mathbb{R}^n, \mathcal{U})$, where $\mathcal{U}$ is the maximal atlas containing $(\mathrm{Id}, \mathbb{R}^n)$.
(2) $(\mathbb{R}, \mathcal{V})$, where $\mathcal{V}$ is the maximal atlas containing the chart $(x \mapsto x^3, \mathbb{R})$. Note that this is different from the standard structure! (Why?)
(3) $\mathbb{S}^n$ equipped with the stereographic projection maps $P_{j,\pm}$ for $j = 1, \ldots, n+1$:

$$\begin{aligned} P_{j,\pm} &: U_{j,\pm} \to \mathbb{R}^n, \\ U_{j,\pm} &= \{v \in \mathbb{R}^{n+1} \mid |v| = 1, v \neq \pm e_j\} \\ P_{j,\pm}(x) &= \Pi_j \left(\pm e_j + \frac{1}{\pm 1 - x^j}(x \mp e_j) \right), \end{aligned}$$

where e_j is the j-th standard basis vector in $\mathbb{R}^{n+1}$ and Π_j denotes the projection $\mathbb{R}^{n+1} \to \mathbb{R}^n$ given by dropping the j-th coordinate. (Draw a picture.)

(4) Products of manifolds have charts given by product charts.
(5) Any open subset of a manifold is again a manifold (with charts given by restriction).

Definition 19. $f : M \to N$ is smooth (or C^k, etc) if for every chart (x, U) ($x(U) \subset \mathbb{R}^n$) on M and (y, V) ($y(V) \subset \mathbb{R}^m$)on N, the composition $y \circ f \circ x^{-1}|_{x(f^{-1}(V))}$ is smooth.

$$\begin{array}{ccc} M & \xrightarrow{f} & N \\ {\scriptstyle x^{-1}}\uparrow & & \downarrow{\scriptstyle y} \\ x(U) & \longrightarrow & y(V) \end{array}$$

A function $f : M \to N$ is a diffeomorphism if it is a smooth homeomorphism so that f^{-1} is also smooth.

Note that strictly speaking, we consider the smoothness of $y \circ f \circ x^{-1}|_{x(f^{-1}(V))}$; we'll drop the domain restriction from our notation in the future, though. That's not to say that it isn't there, we will just leave it implicit.

We also note here that all local notions that are invariant under diffeomorphism can be expressed on manifolds. A big example that we will occasionally use, sometimes without mention, is the property of a subset $A \subset M$ having measure zero.

Properties of differentiable functions:

(1) A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable in the sense of smooth manifold maps if and only if it is differentiable in the standard sense.
(2) $f : M \to \mathbb{R}^k$ is differentiable if and only if each component $f^j : M \to \mathbb{R}$ is differentiable.
(3) $f : M \to N$ is differentiable if and only if each coordinate function $y^i \circ f : M \to \mathbb{R}$ is differentiable for coordinates on N.
(4) If (x, U) is a chart, then $x : U \to x(U)$ is a diffeomorphism.

For $f : \mathbb{R}^n \to \mathbb{R}$, let's use the following notation (at least for now) for the partial derivative at a in the j-th coordinate direction:

$$D_j f(a) = \lim_{h \to 0} \frac{f(a^1, \ldots, a^{j-1}, a^j + h, a^{j+1}, \ldots, a^n) - f(a)}{h}.$$

Now suppose $f : M \to \mathbb{R}$ and (x, U) is a chart on M. For $p \in U$, we define

$$\frac{\partial f}{\partial x^j}(p) = \frac{\partial f}{\partial x^j}|_p = D_j \left(f \circ x^{-1}\right)(x(p)).$$

Proposition 20. *If (x, U) and (y, V) are charts on M and $f : M \to \mathbb{R}$ is differentiable, then*

$$\frac{\partial f}{\partial y^i} = \sum_{j=1}^{n} \frac{\partial f}{\partial x^j} \frac{\partial x^j}{\partial y^i}.$$

Proof. At a point p in V, we have

$$\frac{\partial f}{\partial y^i}(p) = D_i(f \circ y^{-1})(y(p)) = D_i \left((f \circ x^{-1}) \circ (x \circ y^{-1})\right)(y(p)).$$

Using the chain rule and unwinding the meanings of the expressions finishes the proof. □

We'll fill this space soon with some words about manifolds with boundary.

3.2. **Partitions of unity.** In your first homework, you constructed compactly supported smooth functions on $\mathbb{R}$ that were identically one on a neighborhood of the origin. By taking products, rescaling, and translating, given any two open cubes C_1 and C_2 with $C_1 \subset \overline{C_1} \subset C_2$, you can construct smooth functions on $\mathbb{R}^n$ that are identically one on C_1 and supported in C_2. Similarly, given any $0 < r_1 < r_2$, by using your one-variable function, you can use $|x - p|$ as an argument and construct a smooth function that is identically one on $B(p, r_1)$, bounded between 0 and 1, and supported in $B(p, r_2)$. We'll call such functions smooth bump functions.

Definition 21 (From Lee's book)**.** Suppose $\mathcal{U} = (U_\alpha)_{\alpha \in A}$ is an arbitrary open cover of a topological space M. A *partition of unity subordinate to* $\mathcal{U}$ is an indexed family $\{\psi_\alpha\}_{\alpha\in A}$ of continuous functions $\psi_\alpha : M \to \mathbb{R}$ so that

(a) $0 \leq \psi_\alpha(x) \leq 1$ for all $\alpha \in A$ and $x \in M$,
(b) $\operatorname{supp} \psi_\alpha \subset U_\alpha$ for all $\alpha \in A$,
(c) The family of supports $\{\operatorname{supp} \psi_\alpha\}_{\alpha\in A}$ is locally finite, meaning that every point has a neighborhood that intersects $\operatorname{supp} \psi_\alpha$ for only finitely many α.
(d) $\sum_{\alpha\in A} \psi_\alpha(x) = 1$ for all $x \in M$.

On a smooth manifold M, a smooth partition of unity is one for which each function is smooth.

The existence of partitions of unity is essentially why we wanted the second-countability hypothesis in our definition of a manifold. To construct them, we first need to know that we can fill out open covers with sets that look like coordinate balls in $\mathbb{R}^n$.

There are various relaxations of the hypotheses of this useful lemma available, but we'll state it right now in the form that we'll use.

Lemma 22 (Adapted from Lee's Theorem 1.15)**.** *Given a smooth manifold M and an open cover $\mathcal{U}$ of M, there is a countable, locally finite open refinement of $\mathcal{U}$ consisting of open sets B_i diffeomorphic to coordinate balls in $\mathbb{R}^n$. Moreover, the (closed) cover $\{\overline{B_i}\}$ is also locally finite.*

Proof. We first observe that the collection $\mathcal{B}$ of all open sets in M diffeomorphic to coordinate balls in $\mathbb{R}^n$ is a basis[1] for the topology of M.

If M is compact, we can find a finite cover of M by charts with domains diffeomorphic to coordinate balls. (Convince yourself that such charts exist, then take one for each point, then take a finite subcover.)

We may therefore assume that M is non-compact. We let K_j, $j = 1, 2, \dots$ be compact sets so that $K_i \subset \operatorname{Int} K_{i+1}$ and $\cup_i K_i = M$. (The existence of such a family follows from the fact that M is Hausdorff and second-countable.) For each j, let $C_j = K_{j+1} \setminus \operatorname{Int} K_j$ and $V_j = \operatorname{Int} K_{j+2} \setminus K_{j-1}$ (where $K_j = \emptyset$ if $j < 1$), so that $C_j \subset V_j$, C_J is compact, and V_j is open. For each $x \in C_j$, there is some $U_x \in \mathcal{U}$ containing x, and, since $\mathcal{B}$ is a basis, some $B_x \in \mathcal{B}$ with $x \in B_x \subset U_x \cap V_j$.

The collection of all such sets as x ranges over C_j is an open cover of C_j and so has a finite subcover. The union of all of these finite subcovers as $j = 1, 2, \dots$ is a countable open

[1] Recall that this means that, for any point p and open set U containing p, there is some $B \in \mathcal{B}$ with $p \in B \subset U$.

subcover of M that refines $\mathcal{U}$. Because the subcover of C_j consists of sets contained in V_j, and $V_j \cap V_i = \emptyset$ if $|j - j'| > 2$, the resulting cover is locally finite. Because the closures are also in V_j, the cover by the closures remains locally finite. □

Theorem 23 (Lee, Theorem 2.23)**.** *Suppose M is a smooth manifold with or without boundary and $\mathcal{U} = \{U_\alpha\}_{\alpha \in A}$ is an open cover of M. There exists a smooth partition of unity subordinate to $\mathcal{U}$.*

Proof. We'll just prove the boundary-less case now; the case with boundary is similar. (EXERCISE!) By Lemma 22, $\mathcal{U}$ has a countable, locally finite refinement $\{B_i\}$ consisting of neighborhoods B_i diffeomorphic to coordinate balls in $\mathbb{R}^n$ so that $\{\overline{B_i}\}$ is also locally finite.

For each i, we can find some slightly larger B_i' diffeomorphic to a coordinate ball in $\mathbb{R}^n$ so that $\overline{B_i} \subset B_i'$ as well as a coordinate map $x_i : B_i' \to \mathbb{R}^n$ so that $x_i(B_i) = B(0, r_i)$ and $x_i(B_i') = B(0, r_i')$ for some $0 < r_i < r_i'$. Let $H_i : \mathbb{R}^n \to \mathbb{R}$ denote a smooth function that is positive in $B(0, r_i)$ and zero elsewhere, and then define

$$f_i = \begin{cases} H_i \circ x_i & \text{on } B_i' \\ 0 & \text{on } M \setminus \overline{B_i} \end{cases}.$$

In the overlap $B_i' \setminus \overline{B_i}$, the two definitions both yield the zero function, so f_i is a well-defined smooth function on M.

Define $f : M \to \mathbb{R}$ by $f(x) = \sum_i f_i(x)$. As the cover $\overline{B_i}$ is locally fintie, this sum has only finitely many nonzero terms in a neighborhood of each point and therefore defines a smooth function. Each f_i is nonnegative everywhere and positive on B_i, and every point of M is in some B_i, so $f(x) > 0$ everywhere on M. The functions $g_i = f_i(x)/f(x)$ are thus smooth and so $0 \leq g_i(x) \leq 1$ and $\sum_i g_i(x) = 1$. Note that each g_i has support $\overline{B_i}$, which is contained in U_α for some α.

After grouping together terms and re-indexing (EXERCISE), we finish the proof. □

4. Tangent bundles

4.1. **Some words about vector bundles.** Suppose B is an n-dimensional manifold and E is an $(n + k)$-dimensional manifold and there is a map $\pi : E \to B$. We want to define the notion of a vector bundle, which is where we think of each point in a manifold B (for "base") as having attached a finite dimensional vector space attached to it. Even though we'll demand that all of these vector spaces be abstractly isomorphic (i.e., they'll have the same dimension) you should really think of them as *distinct* vector spaces since any isomorphism you come up with will typically depend on a lot of choices.

Definition 24. We say E is a vector bundle over B if

(1) π is surjective,
(2) $\pi^{-1}(p)$ is a k-dimensional vector space for each $p \in B$, and
(3) for each $p \in B$, there is chart (x, U) around p in B and a diffeomorphism $\varphi : \pi^{-1}(U) \to x(U) \times \mathbb{R}^n$ that restricts to be a vector space isomorphism on each fiber, i.e., $\varphi : \pi^{-1}(p) \to \{x(p)\} \times \mathbb{R}^n$ is a a vector space isomorphism for each $p \in U$.

In the above, you think of the space $\{x(p)\} \times \mathbb{R}^n$ as a vector space equipped with the addition and scaling laws of

$$\begin{aligned}(x(p), v) + (x(p), w) &= (x(p), v + w),\\ c(x(p), v) &= (x(p), cv).\end{aligned}$$

Definition 25. A bundle map of two vector bundles is one that preserves the fibers and is linear on each fiber.

You can think of a bundle map as secretly being two maps. In other words, if E_1 is a vector bundle over B_1, E_2 is a vector bundle over B_2, and $f : E_1 \to E_2$ is a bundle map, it induces a map $f_B : B_1 \to B_2$ making the following diagram commute:

$$\begin{tikzcd} E_1 \arrow[r, "f"] \arrow[d, "\pi_1"] & E_2 \arrow[d, "\pi_2"] \\ B_1 \arrow[r, "f_B"] & B_2 \end{tikzcd}$$

Maybe put in a discussion of orientation here in the future? (For now I'll stick a bit closer to Lee's treatment.)

4.2. **The tangent bundle of $\mathbb{R}^n$.** Recall that if $f : \mathbb{R}^n \to \mathbb{R}^m$ is smooth (or, indeed, C^1), then

$$(Df)_p = f'(p) = \left(\frac{\partial f^i}{\partial x^j}(p)\right)$$

is an $m \times n$ matrix and, for $v \in \mathbb{R}^n$,

$$f(p + \epsilon v) = f(p) + \epsilon (Df)_p v + \wr(\epsilon).$$

(If f is smooth, $\wr(\epsilon)$ is in fact $\mathcal{O}(\epsilon^2)$.)

We'd like to simultaneously keep track of the *two* leading terms. For an open subset $U \subset \mathbb{R}^n$, let's say that the tangent bundle of U is given by

$$TU = U \times \mathbb{R}^n,$$

where we think of the first factor as encoding the position p and the second factor encoding the vector v. $T\mathbb{R}^n$ is equipped with a projection

$$\begin{tikzcd} TU \arrow[d, "\pi"] \\ U \end{tikzcd}$$

where $\pi(p, v) = p$. For each p, $\pi^{-1}(p)$ is a vector space. In other words, TU is a vector bundle over U. We refer to the fiber $\{p\} \times \mathbb{R}^n = \pi^{-1}(p)$ as the tangent space of U at p, or T_pU.

If $U \subset \mathbb{R}^n$ and $V \subset \mathbb{R}^m$ are open and $f : \mathbb{R}^n \to \mathbb{R}^m$ is smooth, we then get, for each $p \in U$, $Df_p : \mathbb{R}^n \to \mathbb{R}^m$; putting these together at each point yields a map

$$Df : TU \to TV, \quad (p, v) \mapsto (f(p), Df_p(v)),$$

so that the following diagram commutes

$$\begin{tikzcd} TU \arrow[r, "Df"] \arrow[d, "\pi"] & TV \arrow[d, "\pi"] \\ U \arrow[r, "f"] & V \end{tikzcd}$$

and Df is linear on the fibers of π. In other words, if $f : \mathbb{R}^n \to \mathbb{R}^m$ is smooth, it induces a bundle map $Df : T\mathbb{R}^n \to T\mathbb{R}^m$.

One upshot of this is the following: if you change coordinates on $\mathbb{R}^n$, you automatically induce an isomorphism on the fibers of $T\mathbb{R}^n$. Indeed, if $U, V \subset \mathbb{R}^n$ are open and $f : U \to V$ is a diffeomorphism, then (because f is smoothly invertible), $Df_p : T_pU \to T_{f(p)}V$ must be an isomorphism. (What this will mean on manifolds: once you pick coordinates in a neighborhood, you automatically pick a basis for the tangent space at each point in that neighborhood and therefore also induce coordinates on the tangent bundle over that neighborhood.)

4.3. **How to think of the tangent spaces.** Given a vector $v \in T_pU$, you can associate to it a linear map $C^\infty(U) \to \mathbb{R}$ given by taking the directional derivative at p in direction v. In other words, you can associate to $v \in T_pU$ the map $D_v|_p$ (Lee's notation) given by

$$D_v|_p f = D_v f(p) = \frac{d}{dt}|_{t=0} f(p + tv).$$

If we write v in terms of its components in $\{p\} \times \mathbb{R}^n$ as

$$v = \begin{pmatrix} v^1 \\ v^2 \\ \dots \\ v^n \end{pmatrix},$$

then the chain rule tells us that

$$D_v|_p f = \sum_{j=1}^n v^j \frac{\partial}{\partial x^j} f(p).$$

In other words, we can think of the identification

$$v \mapsto \sum_{j=1}^n v^j \frac{\partial}{\partial x^j}.$$

Definition 26. Following Lee, if $p \in U$, we define a *derivation at* p to be a map $w : C^\infty(U) \to \mathbb{R}$ that is linear over $\mathbb{R}$ that satisfies

$$w(fg) = f(p)w(g) + g(p)w(f)$$

for all $f, g \in C^\infty(U)$.

Let $\mathcal{D}_p$ denote the set of derivations at p.

It is trivial to check that $\mathcal{D}_p$ is a vector space. The main result of this section is that the identification $v \mapsto D_v|_p$ of elements of T_pU with derivations at p is in fact a natural vector space isomorphism.

Lemma 27 (Lee, Lemma 3.1)**.** *Suppose $p \in U \subset \mathbb{R}^n$, $w \in \mathcal{D}_p$, and $f, g : U \to \mathbb{R}$ are smooth.*

(1) If f is constant, then $w(f) = 0$.

(2) *If* $f(p) = g(p) = 0$, *then* $w(fg) = 0$.

Proof. For the first part, it suffices to prove the result for $f : U \to \mathbb{R}$ so that $f \equiv 1$ and then linearity proves it for other constants.

$$w(f) = w(f^2) = 2f(p)w(f) = 2w(f),$$

so $w(f) = 0$.

For the second part, we observe that

$$w(fg) = f(p)w(g) + g(p)w(f) = 0.$$

□

Proposition 28 (Lee, Proposition 3.2)**.** *Let* $p \in U \subset \mathbb{R}^n$.

(1) *For each vector* $v \in T_pU = \{p\} \times \mathbb{R}^n$, *the map* $D_v|_p : C^\infty \to \mathbb{R}$ *is a derivation at* p.
(2) *The map* $v \mapsto D_v|_p$ *is an isomorphism* $T_pU \to \mathcal{D}_p$.

Proof. The first part follows immediately from the product rule. For the second, we note that the map $v \mapsto D_v|_p$ is linear. Applying it to the coordinate functions x^j shows that it is injective, as

$$D_v|_p(x^j) = \sum_{k=1}^{n} v^k \frac{\partial}{\partial x^k}(x^j) = v^j.$$

To prove surjectivity, let $w \in \mathcal{D}_p$. Let $v^j = w(x^j)$ and let $v = \sum_{j=1}^n v^j e_j$, where $e_j \in \{p\} \times \mathbb{R}^n$ is the j-th standard basis vector. We claim that $D_v|_p = w$.

To see this, we appeal to Taylor's theorem, which allows us to write any $f \in C^\infty(U)$ as

$$f(x) = f(p) + \sum_{j=1}^{n} \frac{\partial}{\partial x^j} f(p)(x^j - p^j) + \sum_{i,j=1}^{n} a_{ij}(x)(x^i - p^i)(x^j - p^j).$$

Applying w to each term and using the previous lemma shows that

$$w(f) = 0 + \sum_{j=1}^{n} \frac{\partial f}{\partial x^j}(p) w(x^j) + 0 = \sum_{j=1}^{n} v^j \frac{\partial f}{\partial x^j}(p) = D_v|_p f.$$

□

A consequence of the previous proposition is that the derivations $\frac{\partial}{\partial x^j}$, $j = 1, \dots, n$ form a basis for $\mathcal{D}_p$.

4.4. **The tangent space on a manifold and the differential of a smooth map.** In the interest of (hopefully) minimizing confusion, let's stick a bit closer to Lee's text for now. Let M be a smooth manifold and $p \in M$. We way that a linear (over $\mathbb{R}$) map $v : C^\infty(M) \to \mathbb{R}$ is a *derivation at* p if it satisfies the Leibniz rule:

$$v(fg) = f(p)v(g) + g(p)v(f).$$

The set of all such derivations at p is called the *tangent space to* M *at* p and is denoted T_pM.

The following lemma is provided in exactly the same way as for $\mathcal{D}_p$ on $\mathbb{R}^n$.

Lemma 29. *Let* M *be a smooth manifold,* $p \in M$, *and* $v \in T_pM$. *Suppose* $f, g \in C^\infty(M)$.

(1) *If* f *is a constant function, then* $v(f) = 0$.
(2) *If* $f(p) = g(p) = 0$, *then* $v(fg) = 0$.

We start by observing that derivations (tangent vectors) act locally:

Lemma 30 (Lee, Proposition 3.8)**.** *Suppose M is a smooth manifold, $p \in M$ and $v \in T_pM$. If $f, g \in C^\infty(M)$ agree in a neighborhood of p then $vf = vg$.*

Proof. Let $h = f - g$, so h vanishes in a neighborhood of p. We may find $\psi \in C^\infty(M)$ so that $\psi \equiv 1$ in a neighborhood of p and ψ vanishes on the support of h. In particular, we have that $(1-\psi)h = h$, so that

$$v(h) = v\left((1-\psi)h\right) = h(p)v(1-\psi) + (1-\psi)(p)v(h) = 0$$

because both h and $1-\psi$ vanish at p. As v is linear, we conclude $vf = vg$. □

If M and N are smooth manifolds and $F : M \to N$ is smooth, we now aim to define the differential DF_p of F at p. This is a map $T_pM \to T_{F(p)}N$, so, given $v \in T_pM$, we need to define a derivation at $F(p)$. We let $DF_p(v)$ denote the derivation at $F(p)$ that acts by

$$DF_p(v)(f) = v\left(f \circ F\right)$$

for all $f \in C^\infty(N)$. (This makes sense because $f \circ F \in C^\infty(M)$ and v is a derivation on M at p.) It is a derivation because, given $f, g \in C^\infty(N)$, we have

$$\begin{aligned} DF_p(v)(fg) &= v\left((fg)\circ F\right) = v\left((f\circ F)\cdot(g\circ F)\right) \\ &= f(F(p))v(g\circ F) + g(F(p))v(f\circ F) = f(F(p))DF_p(v)(g) + g(F(p))DF_p(v)(f). \end{aligned}$$

The following proposition is straightforward.

Proposition 31 (Lee, Proposition 3.6)**.** *Let M, N, and P be smooth manifolds, $F : M \to N$ and $G : N \to P$ be smooth, and let $p \in M$.*

(1) $DF_p : T_pM \to T_{F(p)}N$ is linear.
(2) $D(G\circ F)_p : T_pM \to T_{G(F(p))}P = DG_{F(p)} \circ DF_p$.
(3) $D(\mathrm{Id}_M)_p = \mathrm{Id}_{T_pM}$.
(4) If $F : M \to N$ is a diffeomorphism, then $DF_p : T_pM \to T_{F(p)}N$ is an isomorphism and $(DF_p)^{-1} = D\left(F^{-1}\right)_{F(p)}$.

Proof. All are straightforward and the third and fourth follow from the first two, so let's prove just the second one. Suppose $v \in T_pM$ and $f : P \to \mathbb{R}$ is C^∞. We have

$$\begin{aligned} D(G\circ F)_p(v)(f) &= v\left(f \circ (G\circ F)\right) = v\left((f\circ G)\circ F\right) \\ &= DF_p(v)\left(f\circ G\right) = DG_{F(p)}\left(DF_p(v)\right)(f), \end{aligned}$$

as desired. □

We'll use the following proposition to localize; for an open subset $U \subset M$, it allows us to canonically identify tangent vectors $v \in T_pU$ with those in T_pM.

Proposition 32 (Lee, Proposition 3.9)**.** *Suppose M is a smooth manifold, $U \subset M$ is open and $p \in U$. If $\iota : U \to M$ denotes the inclusion map $\iota(p) = p$, then $D\iota_p : T_pU \to T_pM$ is an isomorphism.*

Proof. We know it is linear, so it suffices to show that it is injective and surjective.

We first consider injectivity. If$v \in T_pU$ satisfies $D\iota_p(v) = 0$, we claim $v = 0$. Indeed, suppose $f \in C^\infty(U)$. We take $\chi \in C^\infty(M)$ so that $\chi \equiv 1$ in a neighborhood of p and

$\operatorname{supp}\chi \subset U$. Then χf and f agreee on an open neighborhood of p, so $v(f) = v(\chi f)$. Now, we may regard χf as an element of $C^\infty(M)$ (after extending by 0), so

$$0 = D\iota_p(v)(\chi f) = v((\chi f) \circ \iota) = v(\chi f),$$

so $v(f) = 0$ and thus $v = 0$.

For surjectivity, we let $v \in T_pM$ be arbitrary and take χ as in the previous paragraph. We define $w \in T_pU$ by

$$wf = v(\chi f),$$

where we regard χf as a function on M. It is straightforward to check that w is a derivation. Now, for any $g \in C^\infty(M)$, we have

$$D\iota_p(w)(g) = w\left(g \circ \iota\right) = v\left(\chi \cdot (g \circ \iota)\right) = v(\chi g) = v(g),$$

with the last equality holding by Lemma 30. $\square$

If M is n-dimensional, we can consider an open set around p diffeomorphic to $\mathbb{R}^n$ immediately conclude that the dimension of T_pM is n.

As we keep saying in class, choosing coordinates (by using a chart) near p in a manifold M automatically selects a basis for T_pM. Concretely, let (φ, U) denote a chart[2] around $p \in M$. By combining the propositions above, we see that $D\varphi_p : T_pM \to T_{\varphi(p)}\mathbb{R}^n$ is an isomorphism. As $\frac{\partial}{\partial x^1}|_{\varphi(p)}, \dots, \frac{\partial}{\partial x^n}|_{\varphi(p)}$ form a basis for $T_p\mathbb{R}^n$ (NOTE: here we are already using the identification above of the tangent space to $\mathbb{R}^n$ and the space of derivations.), their preimages under $D\varphi_p$ form a basis for T_pM. For now and forever, we'll use notation conflating them, i.e., we'll say that $\frac{\partial}{\partial x^j}|_p$ denotes the preimage of $\frac{\partial}{\partial x^j}|_{\varphi(p)}$, i.e.,

$$\frac{\partial}{\partial x^j}|_p = (D\varphi_p)^{-1}\left(\frac{\partial}{\partial x^j}|_{\varphi(p)}\right) = D(\varphi^{-1})_{\varphi(p)}\left(\frac{\partial}{\partial x^j}|_{\varphi(p)}\right).$$

You should check from the definition that $\frac{\partial}{\partial x^j}|_p$ is the derivation that sends f to the j-th partial derivative of f in the coordinate system given by the chart.

Given a chart (x, U), we can then write a tangent vector $v \in T_pM$ in terms of this basis. Writing

$$v = \sum_{j=1}^n v^j \frac{\partial}{\partial x^j}|_p,$$

what are the coefficients of v? We follow our nose as we did in $\mathbb{R}^n$, and observe that, regarding x^k as a smooth function on U,

$$\frac{\partial}{\partial x^j}(x^k) = \begin{cases} 1 & j = k \\ 0 & j \neq k \end{cases},$$

so that

$$v^j = v(x^j).$$

What does the differential look like? Suppose $F : M \to N$ is a smooth map and that (φ, U) is a chart around p in M and that (ψ, V) is a chart around $F(p)$ in N. We know from

[2]For this brief discussion, we depart from our convention of using letters like x for charts.

our work in $\mathbb{R}^n$ that if G is a smooth function from an open subset of $\mathbb{R}^n$ to an open subset of $\mathbb{R}^m$, then the total derivative of G at a point $q \in \mathbb{R}^n$ is given by the matrix

$$\begin{pmatrix} \frac{\partial G^1}{\partial x^1} & \cdots & \frac{\partial G^1}{\partial x^n} \\ \vdots & \ddots & \vdots \\ \frac{\partial G^m}{\partial^1} & \cdots & \frac{\partial G^m}{\partial x^n} \end{pmatrix}.$$

Now, letting $G = \psi \circ F \circ \varphi^{-1}$, we compute

$$\begin{aligned} DF_p\left(\frac{\partial}{\partial x^k}|_p\right) &= DF_p\left(D(\varphi^{-1})_{\varphi(p)}\left(\frac{\partial}{\partial x^k}|_{\varphi(p)}\right)\right) \\ &= D(\psi^{-1})_{\psi(F(p))}\left(DG_{\varphi(p)}\left(\frac{\partial}{\partial x^k}|_{\varphi(p)}\right)\right) \\ &= D(\psi^{-1})_{\psi(F(p))}\left(\sum_{j=1}^m \frac{\partial G^j}{\partial x^k}(\varphi(p))\frac{\partial}{\partial y^j}|_{\psi(F(p))}\right) \\ &= \sum_{j=1}^m \frac{\partial G^j}{\partial x^k}(\varphi(p))\frac{\partial}{\partial y^j}|_{F(p)}. \end{aligned}$$

Now that we have the differential in coordinates, what happens when we change coordinates? Suppose (x, U) and $(\tilde{x}, U)$ are two charts around p. (By taking the intersection of the charts, we can assume x and $\tilde{x}$ have the same domain.) Using the above computation, if we write $v \in T_pM$ as

$$v = \sum_{j=1}^n v^j \frac{\partial}{\partial x^j}|_p = \sum_{j=1}^n \tilde{v}^j \frac{\partial}{\partial \tilde{x}}|_p,$$

we have that the components $\tilde{v}$ satisfy[3]

$$\tilde{v}^k = \sum_{j=1}^n \frac{\partial \tilde{x}^k}{\partial x^j} v^j.$$

We'll finish this subsection by noting that the tangent space of a vector space at a point is canonically isomorphic to the vector space itself. (In particular, you now know some more tangent spaces – $\mathrm{GL}(n)$ is an open subset of $n \times n$ matrices, so its tangent space at every point is the space of $n \times n$ matrices.)

Proposition 33 (Lee, Proposition 3.13)**.** *Suppose V is a finite dimensional vector space with its standard smooth structure. For each $a \in V$, the map $v \mapsto D_v|_a$ is a canonical isomorphism from V to T_aV in the sense that for any linear transformation $L : V \to W$, the following diagram commutes:*

$$\begin{tikzcd} V \arrow[r, "\cong"] \arrow[d, "L"] & T_aV \arrow[d, "DL_a"] \\ W \arrow[r, "\cong"] & T_{L(a)}W \end{tikzcd}$$

[3]This is yet another of the abuses/conveniences of notation that are extremely common: we are using $\tilde{x}^j$ to denote the coordinate functions as functions on $x(U)$ and we are using x to denote the points in $x(U)$.

Proof. Once we choose a basis, the same argument we gave in $\mathbb{R}^n$ shows that $v \mapsto D_v|_a$ is an isomorphism.

If $L : V \to W$ is a linear map and $f : T_{L(a)}W \to \mathbb{R}$ is smooth, we compute

$$\begin{aligned} DL_a\left(D_v|_a\right)(f) = D_v|_a(f \circ L) &= \frac{d}{dt}\left((f \circ L)(a + tv)\right)|_{t=0} \\ &= \frac{d}{dt} f(La + tLv)|_{t=0} = D_{Lv}|_{La}, \end{aligned}$$

as desired. ☐

4.5. **The tangent bundle on a manifold.** At last we build the tangent bundle TM on a manifold. As a point set, we take the disjoint union of T_pM as p ranges over M. It is equipped with the natural projection map $\pi : TM \to M$ given by $\pi(p, v) = p$. Note that our earlier discussion of open subsets of $\mathbb{R}^n$ gives a manifold structure to TU for all $U \subset \mathbb{R}^n$ open.

Proposition 34 (Lee, Proposition 3.18)**.** *For any smooth n-manifold M, the tangent bundle TM has a natural topology and smooth structure that make it into a smooth $2n$-manifold. With respect to this structure, the map $\pi : TM \to M$ is smooth.*

Proof. We aren't going to prove this here; I encourage you to read it. I will tell you what the maps are that give the smooth structure. Given a chart (x, U) on M, we define a chart $(\hat{x}, \pi^{-1}(U)$ for TM. Indeed, for $p \in M$ and $v \in T_pM$, we define

$$\hat{x}(p, v) = \left(x(p), (Dx)_p(v)\right).$$

(In other words, if $v = \sum_j v^j \frac{\partial}{\partial x^j}|_p$, $\hat{x}(p, v) = (x^1(p), \dots, x^n(p), v^1, \dots, v^n)$.)

There are a few things you need to check about these maps; the main one is that they are all C^∞-related. ☐

Given a smooth map $F : M \to N$, we then get the "global differential" $DF : TM \to TN$ given by $(p, v) \mapsto (F(p), DF_p(v))$. You need to check this, but it is not so bad:

Proposition 35. *The differential of a smooth map is smooth and satisfies*

1. $D(F \circ G) = DF \circ DG$,
2. $D(\mathrm{Id}_M) = \mathrm{Id}_{TM}$, *and*
3. *if F is a diffeomorphism then $DF : TM \to TN$ is a diffeomorphism and $(DF)^{-1} = D(F^{-1})$.*

Notation alert: Sometimes the "global differential" is called the "pushforward" and is denoted by F_*. This terminology arises because F_* "pushes forward" vectors from TM to TN.

5. Cotangent bundles

5.1. **Vector space duals.** Suppose V is a finite dimensional vector space. We say ω is a *covector* on V if $\omega : V \to \mathbb{R}$ is linear. The space of all covectors on V is a vector space and is called V^*, the *dual vector space to V*.

The proof of the following proposition has a useful construction in it, called the *dual basis*.

Proposition 36. *If V is an n-dimensional vector space, then V^* is also n-dimensional.*

Proof. Let $e_1, \dots, e_n$ denote a basis for V. Let $\epsilon^1, \dots, \epsilon^n : V \to \mathbb{R}$ be linear and satisfy

$$\epsilon^i(e_j) = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}.$$

(Because the e_j form a basis, the condition on each ϵ^i uniquely defines a linear function $V \to \mathbb{R}$.) We claim the ϵ^i form a basis for V^*.

To see they are linearly independent, consider when

$$c_1\epsilon^1 + \cdots + c_n\epsilon^n = 0.$$

Applying both sides to e_j shows $c_j = 0$, so the ϵ^j are linearly independent.

To see they span the entire space, let $\lambda \in V^*$ and define $c_j = \lambda(e_j)$. We then have

$$\lambda = c_1\epsilon^1 + \cdots + c_n\epsilon^n$$

because both sides agree on all elements of V. □

Concretely, once you fix a basis for your vector space V, you can identify vectors in V with column vectors in $\mathbb{R}^n$. The dual of the space of column vectors in $\mathbb{R}^n$ is the space of row vectors with n components (with evaluation being row-column multiplication). We warn that this identification depends crucially on the choice of basis!

Another consequence of the proposition is that finite dimensional vector spaces are isomorphic to their duals but *in a basis-dependent way.*

Now suppose V, W are finite-dimensional vector spaces and $L : V \to W$ is a linear transformation. The map L induces a map of the dual spaces, denoted L^* (and called the dual map or the transpose or the pullback): for $\lambda \in W^*$,

$$L^* : W^* \to V^*, \quad (L^*\lambda)(v) = \lambda(Lv).$$

Observe that $(L \circ M)^* = M^* \circ L^*$ and that $\mathrm{Id}_V^* = \mathrm{Id}_{V^*}$.

We now turn our attention to the dual of V^*, called the "double dual" of V and denoted V^{**}. First observe that there is a natural (i.e., choice-free) linear transformation $i_V : V \to V^{**}$ taking v to the map "evaluate on v" from V^* to $\mathbb{R}$:

$$i_V(v)(\lambda) = \lambda(v).$$

In general, this is a pretty good map. For finite-dimensional vector spaces, it's a great one:

Proposition 37. *If V is a finite-dimensional vector space, then the map $i_V : V \to V^{**}$ is an isomorphism.*

Proof. Since $\dim V^{**} = \dim V$, it suffices to show that i_V is injective. Suppose $v \in V$ is nonzero, so we can extend it to a basis $\{e_1 = v, e_2, \dots, e_n\}$ of V. Let ϵ^i denote the corresponding dual basis of V^*. We then have that

$$i_V(v)(\epsilon^1) = \epsilon^1(v) = \epsilon^1(e_1) = 1,$$

so that $i_V(v) \neq 0$ and thus i_V is injective. □

A word of warning: Although we can think of V^* as the space of linear functionals on V, the identification of V with V^{**} above also lets us identify V with the space of linear functionals on V^*. We will go back and forth between these points of view at will. BE ON YOUR TOES!

5.2. **The cotangent space.** Given a smooth manifold M and a point p, we define the cotangent space to M at p as the dual of the tangent space:

$$T_p^*M = (T_pM)^* .$$

Aside: If you are more algebraically minded, you could instead take the cotangent bundle to be your primary object and define it as follows. Given $p \in M$, let $\mathcal{I}_p = \{f \in C^\infty(M) \mid f(p) = 0\}$ be the ideal of smooth functions vanishing at p. Let $\mathcal{I}_p^2$ denote the squre of this idea, i.e., the ideal of functions of the form $\sum_{\text{finite}} fg$, where $f, g \in \mathcal{I}_p$. The cotangent space to M at p is then $\mathcal{I}_p/\mathcal{I}_p^2$; one consequence of Taylor's theorem is that this space is n-dimensional if M is. We could then have defined the tangent space to be the dual of this space.

Recall that given a chart (x, U) on M, we get a basis

$$\frac{\partial}{\partial x^1}|_p, \dots, \frac{\partial}{\partial x^n}|_p$$

for T_pM at each $p \in U$. Let $\lambda^1|_p, \dots, \lambda^n|_p$ denote the corresponding dual basis for T_p^*M. (These names will change soon.) Given a covector $\omega \in T_p^*M$, we can write it uniquely as $\omega = \sum_{j=1}^n \omega_j \lambda^j$, where

$$\omega_j = \omega\left(\frac{\partial}{\partial x^j}|_p\right).$$

How do they transform under a change of coordinates? Suppose $\tilde{x}$ is another set of coordinates and $\tilde{\lambda}^j|_p$ is the corresponding basis of T_p^*M. We know from above that

$$\frac{\partial}{\partial x^k}|_p = \sum_{j=1}^n \frac{\partial \tilde{x}^j}{\partial x^k}\frac{\partial}{\partial \tilde{x}^j}|_p.$$

In particular,

$$\begin{aligned}
\tilde{\lambda}^\ell|_p\left(\frac{\partial}{\partial x^k}|_p\right) &= \sum_{j=1}^n \frac{\partial \tilde{x}^j}{\partial x^k}\tilde{\lambda}^\ell|_p\left(\frac{\partial}{\partial \tilde{x}^j}|_p\right) \\
&= \frac{\partial \tilde{x}^\ell}{\partial x^k}(p).
\end{aligned}$$

We may therefore express $\tilde{\lambda}^\ell|_p$ in terms of $\lambda^j|_p$:

$$\tilde{\lambda}^\ell|_p = \sum_{j=1}^n \frac{\partial \tilde{x}^\ell}{\partial x^j}\lambda^j|_p.$$

We now take $\omega \in T_p^*M$. Writing

$$\omega = \sum_{j=1}^n \omega_j \lambda^j|_p = \sum_{j=1}^n \tilde{\omega}_j \tilde{\lambda}^j|_p,$$

we want to relate the coefficients $\tilde{\omega}_j$ and ω_j. Substituting in our expression for $\tilde{\lambda}$ in terms of λ, we get

$$\sum_{j=1}^n \tilde{\omega}_j \frac{\partial \tilde{x}^j}{\partial x^k} = \omega_k.$$

5.3. **The differential of a function.** We want to get an understanding of the dual space of T_pM. What do covectors in T_p^*M look like? In the case of open sets in $\mathbb{R}^n$, if we represent a vector in T_pM by its components as a column vector, then covectors in T_p^*M can be represented as row vectors and the pairing between T_p^*M and T_pM is just row-column multiplication.

What about more generally? We know that vectors in T_pM act as derivations on functions and that T_pM is a vector space. Flipping this around, given a function $f \in C^\infty(M)$, we can view f as giving a linear map $T_pM \to \mathbb{R}$ by

$$v \mapsto v(f).$$

As a linear map $T_pM \to \mathbb{R}$ is nothing but an element of T_p^*M, we obtain a map $C^\infty(M) \to T_p^*M$. We'll denote this map by $f \mapsto df_p$.

What does df_p look like in coordinates? Suppose (x, U) is a coordinate chart around p, so that $\frac{\partial}{\partial x^1}, \dots, \frac{\partial}{\partial x^n}$ is a basis for T_pM. We observe that

$$df_p\left(\frac{\partial}{\partial x^K}\right) = \frac{\partial}{\partial x^k}(f) = \frac{\partial f}{\partial x^k}(p).$$

In terms of the dual basis λ^j to the $\frac{\partial}{\partial x^k}$, we can write

$$df_p = \sum_{j=1}^n \frac{\partial f}{\partial x^j}(p)\lambda^j|_p.$$

If we apply this reasoning to the local coordinate functions x^k, we see that in fact

$$dx_p^k = \sum_{j=1}^n \frac{\partial x^k}{\partial x^j}(p)\lambda^j|_p = \sum_{j=1}^n \delta_j^k \lambda^j|_p,$$

so that in fact

$$\lambda^k|_p = dx_p^k.$$

In particular, the differential is surjective $C^\infty(M) \to T_p^*M$ and so every element of T_p^*M is df_p for some f. Thanks to this identity, now and forever we will use dx^k to denote the dual basis to $\partial/\partial x^j$.

5.4. **The cotangent bundle.** The change of coordinates expressions we obtained in Section 5.3 above allow us to create a vector bundle T^*M over M with fibers T_p^*M. Just as in the case of the tangent bundle, if M is a smooth n-manifold, the cotangent bundle T^*M has a natural topology and smooth structure making it a $2n$-manifold. The projection $\pi : T^*M \to M$ given by $\pi(p, \xi) = p$ is smooth with respect to this structure.

The differential of a smooth function $f : M \to \mathbb{R}$ is a "section" of the cotangent bundle,[4] i.e., we can assemble df_p at every point p to get a map $df : M \to T^*M$ given by $p \mapsto (p, df_p)$.

Given a smooth map $F : M \to N$ and an element $\xi \in T_{F(p)}^*N$, we can "pull back" ξ to an element $F^*\xi \in T_p^*M$; for $v \in T_pM$, this is given by

$$(F^*\xi)(v) = \xi(DF_p(v)) = \xi(F_*v).$$

In terms of differentials, you can compute (and YOU SHOULD), that the pullback is given by

$$F^*(df_{F(p)}) = d(f \circ F)_p.$$

[4] A section of a vector bundle E over M is a smooth map $s : M \to E$ so that $\pi \circ s = \mathrm{Id}_M$.

We also recall that functions "pull back", i.e., given a smooth map $F : M \to N$, we get a linear map $F^* : C^\infty(N) \to C^\infty(M)$ given by $F^* f = f \circ F$. The pullback commutes with the differential, so that given $f \in C^\infty(N)$,

$$F^* df = d\left(F^* f\right).$$

(EXERCISE: Check that this makes sense and is true.)

6. Tensors

The word tensor in geometry often has two (related) meanings: on the one hand it can refer to an element of a general tensor product of (powers of) T_pM and T_p^*M, while on the other, it can refer to sections of the tensor bundle. These two meanings are often conflated (especially in undergraduate physics courses), and we'll eventually end up conflating them, too. Before we do that, though, we'll try to pin down both meanings.

6.1. **Tensor products of vector spaces.** Suppose V and W are finite dimensional vector spaces. Our aim is to define the *tensor product of V and W*, denoted $V \otimes W$. We'll describe it via its universal property and then show that such an object exists. (The universal property and the basis are how you actually work with it.) Part of the idea is that bilinear maps $V \times W \to X$ aren't linear transformations but they shouldn't be so far from it.

Suppose V, W, and X are finite dimensional vector spaces. We say that a map $T : V \times W \to X$ is *multilinear* if, for all $v_1, v_2, v \in V$, $\alpha, \beta \in \mathbb{R}$, and $w_1, w_2, w \in W$, T satisfies

$$\begin{aligned} T(\alpha v_1 + \beta v_2, w) &= \alpha T(v_1, w) + \beta T(v_2, w), \\ T(v, \alpha w_1 + \beta w_2) &= \alpha T(v, w_1) + \beta T(v, w_2). \end{aligned}$$

(An analogous definition applies to multilinear maps $V_1 \times \cdots \times V_r \to X$.)

The *tensor product* $V \otimes W$ of V and W is defined by the universal property that mulitlinear maps $V \times W \to X$ factor through $V \otimes W$. More precisely, $V \otimes W$ is a finite dimensional vector space and comes with a multilinear map $\varphi : V \times W \to V \otimes W$. The map φ has the property that if $T : V \times W \to X$ is a multilinear map, then there is a *unique linear transformation* $L : V \otimes W \to X$ so that $T = L \circ \varphi$, i.e., making the following diagram commute.

$$\begin{tikzcd} V \times W \arrow[r, "\varphi"] \arrow[dr, "T"'] & V \otimes W \arrow[d, dashed, "\exists! L"] \\ & X \end{tikzcd}$$

A quick practice using the universal property: $V \otimes W$ is "unique up to unique isomorphism." In other words, if Y_1 and Y_2 are two finite dimensional vector spaces equipped with bilinear maps $\varphi_i : V \times W \to Y_I$ satisfying the universal property above, then there is a unique isomorphism $L : Y_1 \to Y_2$ with $\varphi_2 = L \circ \varphi_1$.

Why? If we take take $X = Y_1$ and $T = \varphi_1$, then the unique map L here is given by $L = \mathrm{Id}_{Y_1}$. Now taking $X = Y_2$ and $T = \varphi_2$, we get the unique linear transformation $L : Y_1 \to Y_2$ compatible with the two different maps. It's an isomorphism because you can swap the roles of Y_1 and Y_2 to get a map going the other way whose composition must be the identity by the uniqueness of the induced map.

Now that we know $V \otimes W$ is essentially unique, we turn to the existence of such an object. Although there are more general approaches, it is expedient for us to use the finite-dimensionality of V and W to our benefit. Suppose $\{e_1, \ldots, e_n\}$ is a basis for V

and $\{f_1, \ldots, f_m\}$ is a basis for W. We let $V \times W$ denote the (formal) span of the $n \times m$ elements $e_i \otimes f_j$:

$$V \times W = \operatorname{span}_{\mathbb{R}}\{e_i \otimes f_j \mid i = 1, \ldots, n,\ j = 1, \ldots, m\}.$$

The multilinear map $\varphi : V \times W \to V \otimes W$ is then given by its action on basis elements:

$$\varphi\left(\sum_{i=1}^{n} \alpha_i e_i, \sum_{j=1}^{m} \beta_j f_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} \alpha_i \beta_j (e_i \otimes f_j).$$

It is straightforward to check that φ is multilinear. Now, given a multilinear map $T : V \times W \to X$, we define L by its action on the basis $e_i \otimes f_j$ of $V \otimes W$:

$$L(e_i \otimes f_j) = T(e_i, e_j).$$

To check that T factors through φ, we then expand in bases and check that

$$\begin{aligned} T(v, w) = T\left(\sum_{i=1}^{n} \alpha_i e_i, \sum_{j=1}^{m} \beta_j f_j\right) &= \sum_{i=1}^{n}\sum_{j=1}^{m} \alpha_i \beta_j T(e_i, f_j) \\ &= \sum_{i,j} \alpha_i \beta_j L(\varphi(e_i, f_j)) \\ &= L\left(\sum_{i,j} \alpha_i \beta_j \varphi(e_i, f_j)\right) = L \circ \varphi(v, w). \end{aligned}$$

From the construction above, given a basis e_i of V and a basis f_j of W, the elements $\varphi(e_i, f_j)$ (from now on denoted[5] $e_i \otimes f_j$) form a basis for $V \otimes W$. As a result, $\dim V \otimes W = nm$ if $\dim V = n$ and $\dim W = m$.

We'll try to keep our notation consistent with Lee's for the next bit. Let V be a finite dimensional vector space and k a positive integer. We define[6] a *covariant k-tensor* on V to be an element of the k-fold tensor product $V^* \otimes \cdots \otimes V^*$; using the universal property of the tensor product, we typically think of a covariant k-tensor as a multilinear map

$$\alpha : V \times \cdots \times V \to \mathbb{R}.$$

The number k is called the rank of α; a 0-tensor is a real number (i.e., a real-valued function depending multilinearly on no vectors). The vector space of all covariant k-tensors is denoted

$$T^k(V^*) = V^* \otimes \cdots \otimes V^* = (V^*)^{\otimes k}.$$

Examples: $T^1(V^*) = V^*$. A covariant 2-tensor is a multilinear function $V \times V \to \mathbb{R}$ – one example is the dot product. If you think of the determinant as a multilinear function on n vectors, it is a covariant n-tensor.

The space of *contravariant tensors on V of rank k* is the vector space

$$T^k(V) = V \otimes \cdots \otimes V = V^{\otimes k},$$

[5]In fact we will denote $T(v, w) = v \otimes w$.

[6]For historical reasons, covariant and contravariant seem to be switched from what you might expect if you are coming from category theory. The naming reason is that the *basis* of V^* transforms "with the metric", i.e., like the components of a vector, while a basis for V^* transforms "against the metric". As we haven't introduced metrics at this stage, this naming convention might seem a bit bonkers.

so that $T^1(V) = V$ and $T^0 = \mathbb{R}$. Because V is finite dimensional, we can identify it with linear functions on V^* and so we typically think of a contravariant tensor on V of rank k as a multilinear function

$$T^k(V) \cong \{\alpha : V^* \times \cdots \times V^* \to \mathbb{R} \mid \alpha \text{ is multilinear}\}.$$

The space of *mixed tensors of type* (k, ℓ) *on* V is given by

$$T^{(k,\ell)}(V) = V^{\otimes k} \otimes (V^*)^{\otimes \ell}.$$

Observe that

$$\begin{aligned} T^{(0,0)}(V) &= T^0(V) = T^0(V^*) = \mathbb{R}, \\ T^{(k,0)}(V) &= T^k(V), \\ T^{(0,\ell)}(V) &= T^\ell(V^*). \end{aligned}$$

Spivak uses the notation $T^k_\ell(V)$, which I prefer – it reminds you which components to write "up" and which to write "down", but we'll stick with Lee's notation.

Lemma 38. *Let V be a finite dimensional vector space. If $\{e_i\}_{i=1}^n$ is a basis for V and $\{\epsilon^j\}_{j=1}^n$ is the corresponding dual basis for V^*, then*

$$\left\{e_{i_1} \otimes e_{i_2} \otimes \cdots \otimes e_{i_k} \otimes \epsilon^{j_1} \otimes \epsilon^{j_2} \otimes \cdots \otimes \epsilon^{j_\ell} \mid 1 \le i_r, j_s \le n\right\}$$

is a basis for $T^{(k,\ell)}(V)$.

In your homework, you will show that the vector space of linear transformations from V to W can be canonically identified with the vector space $V^* \otimes W$. In particular, the space of $(1,1)$-tensors on V can be identified with endomorphisms of V. There is a natural map $V^* \otimes V \to \mathbb{R}$ sometimes called the evaluation map. It for $\lambda \in V^*$ and $v \in V$, the map is given by $\lambda \otimes v \mapsto \lambda(v)$. It extends linearly to a map $V^* \otimes V \to \mathbb{R}$. Under the identification with endomorphisms of V, this map is the trace. I strongly encourage you to think this through. (What does it look like on a basis of $V^* \otimes V$ and how does this relate to what you have previously called the trace?)

6.2. **Tensors on manifolds.** We use the above to describe the tensor bundles on manifolds. Let M be a smooth manifold. The *bundle of covariant k-tensors on M* is defined by

$$T^k T^* M = \coprod_{p \in M} T^k(T_p^* M).$$

Similarly, the *bundle of contravariant k-tensors on M* is given by

$$T^k TM = \coprod_{p \in M} T^k(T_p M),$$

and the *bundle of mixed tensors of type* (k, ℓ) is given by

$$T^{(k,\ell)} TM = \coprod_{p \in M} T^{(k,\ell)}(T_p M).$$

Lemma 39. *Each of the tensor bundles defined above has a natural structure making it into a smooth vector bundle over M; if M is dimension n, $T(k,\ell)TM$ has rank $n^{k+\ell}$ over M.*

If $T \in T^{(k,\ell)}(T_pM)$, in a chart (x, U) around p we can express T in terms of the basis

$$\left\{ \frac{\partial}{\partial x^{i_1}} \otimes \frac{\partial}{\partial x^{i_2}} \otimes \cdots \otimes \frac{\partial}{\partial x^{i_k}} \otimes dx^{j_1} \otimes dx^{j_2} \otimes \cdots \otimes dx^{j_\ell} \mid 1 \leq i_r, j_s \leq n \right\}$$

for $T^{(k,\ell)}(T_pM)$. It is typically written in terms of its components as

$$T = \sum_{i_r, j_s = 1}^{n} T^{i_1 i_2 \dots i_k}_{j_1 j_2 \dots j_\ell} \frac{\partial}{\partial x^{i_1}} \otimes \frac{\partial}{\partial x^{i_2}} \otimes \cdots \otimes \frac{\partial}{\partial x^{i_k}} \otimes dx^{j_1} \otimes dx^{j_2} \otimes \cdots \otimes dx^{j_\ell}.$$

In particular, you obtain the components by plugging in the corresponding basis vectors for T_pM and T_p^*M:

$$T^{i_1 i_2 \dots i_k}_{j_1 j_2 \dots j_\ell} = T\left(dx^{i_1}, \dots, dx^{i_k}, \frac{\partial}{\partial x^{j_1}}, \dots, \frac{\partial}{\partial x^{j_\ell}} \right).$$

Note that by convention the "vector-like" indices for the components of T are written as superscripts and the "covector-like" indices are written as subscripts. Also note that our summations run over indices appearing twice: once "up" and once "down". A common convention in differential geometry is called the "Einstein summation convention" and dictates that when an index appears up and down in a formula, there is an implicit summation over that index.

If (x, U) and $(\tilde{x}, U)$ are both coordinate systems around the point p, then the components of $T \in T^{(k,\ell)}(T_pM)$ change as follows:

$$\begin{aligned}
\tilde{T}^{i_1 \dots i_k}_{j_1 \dots j_\ell} &= T\left(dx^{i_1}, \dots, dx^{i_k}, \frac{\partial}{\partial x^{j_1}}, \dots, \frac{\partial}{\partial x^{j_\ell}} \right) \\
&= T\left(\sum_{r_1} \frac{\partial \tilde{x}^{i_1}}{\partial x^{r_1}} dx^{r^1}, \dots, \sum_{r_k} \frac{\partial \tilde{x}^{i_k}}{\partial x^{r}} dx^{r_k}, \sum_{s_1} \frac{\partial x^{s_1}}{\partial \tilde{x}^{j_1}} \frac{\partial}{\partial x^{s_1}}, \dots, \sum_{s_\ell} \frac{\partial x^{s_\ell}}{\partial \tilde{x}^{j_\ell}} \frac{\partial}{\partial x^{s_\ell}} \right) \\
&= \sum_{r_\mu, s_\nu = 1}^{n} \frac{\partial \tilde{x}^{i_1}}{\partial x^{r_1}} \cdots \frac{\partial \tilde{x}^{i_\ell}}{\partial x^{r_\ell}} \frac{\partial x^{s_1}}{\partial \tilde{x}^{j_1}} \cdots \frac{\partial x^{s_\ell}}{\partial \tilde{x}^{j_\ell}} T^{r_1 \dots r_k}_{s_1 \dots s_k}.
\end{aligned}$$

In other words, each "up" index comes with a factor of the Jacobian matrix and each "down" index comes with a factor of the inverse Jacobian.

7. Vector fields

We now turn our attention to *sections* of the tangent bundle; these are called *vector fields*.

Suppose M is a smooth manifold. A *vector field on M* is a smooth section of the map $\pi : TM \to M$, i.e., it is a map $X : M \to TM$, often written $p \mapsto X_p$, so that for each $p \in M$, $X_p \in T_pM$. In other words, it has the property that

$$\pi \circ X = \mathrm{Id}_M .$$

A *smooth vector field* is a vector field that is also smooth as a map $M \to TM$. The support of a vector field is the closure of the set $\{p \in M \mid X_p \neq 0\}$. A vector field is compactly supported if its support is compact. We denote by $\mathcal{X}(M)$ the space of smooth vector fields on M.

The following lemma is a useful exercise.

Lemma 40 (Lee, Proposition 8.1)**.** *Suppose M is a smooth manifold and X is any (not necessarily smooth) vector field on M. The vector field X is smooth if and only if, for all charts (x, U) in an atlas of M, the components of X with respect to each chart are smooth functions.*

If A is a subset of M, a *vector field along A* is a continuous map $X : A \to TM$ with $\pi \circ X = \mathrm{Id}_A$. A smooth vector field along A is one that is locally the restriction of a smooth vector field. Just as with smooth functions, we can extend smooth vector fields from closed subsets to all of M; the proof is essentially identical.

If $f, g \in C^\infty(M)$ and $X, Y \in \mathcal{X}(M)$, we can create a new smooth vector field by

$$fX + gY, \quad p \mapsto f(p)X_p + g(p)Y_p.$$

With this action, $\mathcal{X}(M)$ is a module over the ring $C^\infty(M)$.

A word of warning: Even though we can push vectors forward by a smooth map (i.e., given $F : M \to N$ and $v \in T_pM$, we obtain $DF_p(v) \in T_{F(p)}N$), doing this to a vector field does *not* in general give a vector field on N. The reason for this is exactly what you might think: if F is not surjective, we do not have an assignment for points outside of the range of F, while if F is not injective, we possibly have at least two different assignments of a vector to a given point. Care must therefore be taken in general.

We can, however, describe a related notion. Suppose $F : M \to N$, $X \in \mathcal{X}(M)$, and $Y \in \mathcal{X}(N)$. We say that X and Y are *F-related* if, for every $p \in M$, $DF_p(X_p) = Y_{F(p)}$.

Lemma 41 (Lee, Proposition 8.16)**.** *Suppose $F : M \to N$ is smooth, $X \in \mathcal{X}(M)$, and $Y \in \mathcal{X}(N)$. Then X and Y are F-related if and only if for every smooth function f on N,*

$$X(f \circ F) = (Yf) \circ F.$$

Proof. Let $p \in M$ and $f \in C^\infty(N)$. Then

$$X(f \circ F)(p) = X_p(f \circ F) = DF_p(X_p)(f),$$

and

$$(Yf) \circ F(p) = (Yf)(F(p)) = Y_{F(p)}(f),$$

proving the claim. □

If $f : M \to N$ is a diffeomorphism, then for every $X \in \mathcal{X}(M)$, there is a unique smooth vector field $Y \in \mathcal{X}(N)$ that is F-related to X; in this case we call Y the *pushforward* of X and denote it F_*X. In particular, we have

$$(F_*X)_q = DF_{F^{-1}(q)}(X_{F^{-1}(q)}).$$

7.1. **Solving differential equations.** If X is a vector field on $\mathbb{R}^n$, we'd like to know when we can find integral curves tangent to it, i.e., functions $\gamma : (-\epsilon, \epsilon) \to \mathbb{R}$ so that

$$\begin{aligned} \frac{d\gamma}{dt} &= X(\gamma(t)), \\ \gamma(0) &= p. \end{aligned}$$

On manifolds, it's similar: given a vector field X on M and a point $p \in M$, we'd like to find $\gamma : (-\epsilon, \epsilon) \to M$ so that

$$\begin{aligned} \frac{d\gamma}{dt} &= X(\gamma(t)), \\ \gamma(0) &= p. \end{aligned}$$

Here $\frac{d\gamma}{dt}$ can be understood as the derivation $\frac{d}{dt}$ pushed forward by γ, i.e., $D\gamma_t(\frac{d}{dt})$.

Note that by passing to charts, the problem on manifolds is essentially equivalent to the corresponding problem on $\mathbb{R}^n$. Given a vector field X, such a curve is called *an integral curve of X with initial condition p.* The aim of this section is to show that integral curves exist.

Lemma 42 (Contraction mapping principle)**.** *Let (Z, d) be a nonempty complete metric space. If $f : Z \to Z$ be a contraction, i.e., there is some $C < 1$ so that for all $x, y \in Z$,*

$$d(f(x), f(y)) \leq Cd(x, y),$$

then there is a unique $x_\infty \in Z$ so that $f(x_\infty) = x_\infty$.

Proof. Note that f is continuous as it is Lipschitz. Pick some $x_0 \in Z$ and define, for $k \geq 0$, $x_{k+1} = f(x_k)$. Note that for $n \geq 1$,

$$d(x_n, x_{n+1}) \leq Cd(x_{n-1}, x_n) \leq C^n d(x_0, x_1),$$

so that

$$\begin{aligned} d(x_n, x_{n+k}) &\leq \sum_{j=1}^{k} d(x_{n+j-1}, x_{n+j}) \\ &\leq \left(C^n + \cdots + C^{n+k-1}\right) d(x_0, x_1) \\ &\leq \frac{C^n}{1-C} d(x_0, x_1), \end{aligned}$$

so that the x_n are a Cauchy sequence and thus converges to some x_∞. As f is continuous, we must have $f(x_\infty) = x_\infty$. □

We'll use this to solve ODEs. The general set up on $\mathbb{R}^n$ is to demand $y : (-\epsilon, \epsilon) \to \mathbb{R}^n$ with

$$\begin{aligned} \frac{dy}{dt} &= f(t, y(t)), \\ y(0) &= y_0. \end{aligned}$$

We demand that f be continuous in t and Lipschitz in y.[7] The strategy is to reformulate the differential equation as an integral one. By the fundamental theorem of calculus, y satisfies the differential equation above if and only if

$$y(t) = y_0 + \int_0^t f(s, y(s))\, ds.$$

We'll use the contraction mapping theorem to find a fixed point of a related integral map, which will lead us to the desired function y.

Theorem 43 (Picard iteration)**.** *Let $f : (t_0 - \epsilon, t_0 + \epsilon) \times U \to \mathbb{R}^n$, where $I \subset \mathbb{R}$ is an open interval and $U \subset \mathbb{R}^n$ is open. Suppose $x_0 \in U$ and take $a > 0$ so that $\overline{B_{2a}(x_0)} \subset U$. Suppose further that*

(1) there is a number L so that $|f| \leq L$ on $I \times \overline{B_{2a}(x_0)}$, and

[7] Recall that f being Lipschitz in y on a set U means that there is some K so that $|f(y_1) - f(y_2)| \leq K\,|y_1 - y_2|$ for all $y_1, y_2 \in U$.

(2) there is a number K so that for all $s, t \in I$ and $x, y \in \overline{B_{2a}(x_0)}$, $|f(t,x) - f(s,y)| \leq K|x-y|$.

Then, for $0 < b < \min(\epsilon, a/L, 1/K)$, for each $x \in \overline{B_a(x_0)}$, there is a unique $\gamma_x : (t_0 - b, t_0 + b) \to U$ so that

$$\begin{aligned} \gamma_x'(t) &= f(t, \gamma_x(t)), \\ \gamma_x(t_0) &= x. \end{aligned}$$

Proof. Fix $x \in \overline{B_a(x_0)}$ and set

$$\mathcal{Z} = \left\{ \gamma : (t_0 - b, t_0 + b) \to \overline{B_{2a}(x_0)} \mid \gamma \text{ is continuous} \right\}.$$

Define a metric (really a norm) on $\mathcal{Z}$ by

$$\|\gamma_1 - \gamma_2\| = \sup_{t \in (t_0 - b, t_0 + b)} |\gamma_1(t) - \gamma_2(t)|.$$

Observe that $\mathcal{M}$ is complete with respect to this metric.

For each $\gamma \in \mathcal{M}$, define $S\gamma : (t_0 - b, t_0 + b) \to \mathbb{R}^n$ by

$$S\gamma(t) = x + \int_{t_0}^t f(s, \gamma(s))\, ds.$$

We claim that $S : \mathcal{M} \to \mathcal{M}$ is a contraction.

First note that if $\gamma \in \mathcal{M}$ then $S\gamma$ is continuous (as it is the integral of a continuous function). We now observe that

$$|S\gamma(t) - x_0| \leq |S\gamma(t) - x| + |x - x_0| \leq \left| \int_{t_0}^t f(s, \gamma(s))\, ds \right| + a \leq \int_{t_0}^{t_0 + b} L\, ds + a \leq bL + a \leq 2a,$$

so that $S\gamma(t) \in \overline{B_{2a}(x_0)}$ and thus $S\gamma \in \mathcal{M}$.

Finally, we note that if $\alpha, \beta \in \mathcal{M}$, we have

$$\begin{aligned} \|S\alpha - S\beta\| &= \sup_{t \in (t_0 - b, t_0 + b)} |S\alpha(t) - S\beta(t)| \\ &= \sup_{t \in (t_0 - b, t_0 + b)} \left| \int_{t_0}^t \left(f(s, \alpha(s)) - f(s, \beta(s)) \right) \right| \\ &\leq \sup_{t \in (t_0 - b, t_0 + b)} \left| \int_{t_0}^t K |\alpha(s) - \beta(s)|\, ds \right| \leq bK \|\alpha - \beta\|. \end{aligned}$$

As $bK < 1$, S is a contraction and thus has a unique fixed point γ_x.

The remaining detail is to show that a fixed point of S must in fact be differentiable and satisfy the differential equation. Because γ and f are continuous in their arguments, $f(t, \gamma(t))$ is a continuous function of t. The fundamental theorem of calculus then implies that

$$\gamma(t) = x + \int_{t_0}^t f(s, \gamma(s))\, ds$$

is a differentiable function and its derivative is

$$\gamma'(t) = f(t, \gamma(t)),$$

as desired. □

We can use this theorem in charts to conclude that our vector fields on manifolds have integral curves. In our special case, $f(t,\gamma) = X(\gamma)$, so the ODE is autonomous. In particular, if $\gamma(t)$ is a solution then so is $\beta(t) = \gamma(t+t_0)$. The solution therefore gives us a flow $\phi_t : U \to \mathbb{R}^n$ where $\phi_t(x) = \gamma_x(t)$. The proof of Picard iteration immediately gives that the flow ϕ_t is continuous. A harder theorem (that we won't bother to prove here, but we'll use) is the following:

Theorem 44. *If f is smooth then so is $\gamma : (t_0 - b, t_0 + b) \times U \to \mathbb{R}^n$.*

Rephrasing the above on manifolds:

Theorem 45. *Let $X \in \mathcal{X}(M)$ be a smooth vector field and let $p \in M$. There is an open set V containing p and an $\epsilon > 0$ so that there is a unique collection of diffeomorphisms $\phi_t : V \to \phi_t(V) \subset M$ for $|t| < \epsilon$ satisfying*

(1) $\phi : (-\epsilon, \epsilon) \times V \to M$ given by $\phi(t,p) = \phi_t(p)$ is C^∞,
(2) if $|s|, |t|, |s+t| < \epsilon$, then $\phi_{s+t} = \phi_s \circ \phi_t$, and
(3) if $q \in V$, then $X(q)$ is the tangent vector at 0 of the curve $t \mapsto \phi_t(q)$.

Proof. Most of the properties follow from the theorems above, but we'll describe the proof of the semigroup property, which essentially follows from uniqueness. Suppose $|s|, |t|, |s+t| < \epsilon$ and $q \in V$. We know that $\phi_{s+t}(q)$ is given by $\gamma_q(t+s)$, where $\gamma_q'(t) = X(\gamma_q(t))$ and $\gamma_q(0) = q$. Similarly, $\phi_t(q) = \gamma_q(t)$. Now $\phi_s(\phi_t(q)) = \gamma_{\phi_t(q)}(s)$. Note that $r \mapsto \gamma_q(r+t)$ is a solution of the differential equation with initial value $\gamma_q(0+t) = \gamma_q(t)$, so we must have that $\gamma_{\phi_t(q)}(r) = \gamma_q(r+t)$ by uniqueness and hence $\phi_{t+s} = \phi_s \circ \phi_t$. □

By using a finite number of open sets we obtain the following:

Theorem 46. *If $X \in \mathcal{X}(M)$ has compact support (e.g., if M is compact), then these diffeomorphisms exist for all $t \in \mathbb{R}$ and all $p \in M$.*

We say that the ϕ_t are the 1-*parameter family of diffeomorphisms* associated to the vector field X. We also remark that $X(q)$ being the tangent vector of $t \mapsto \phi_t(q)$ at $t = 0$ means that

$$(Xf)(q) = \frac{d}{dt}|_{t=0}(f \circ \phi_t(q)) = \lim_{h \to 0} \frac{f(\phi_h(q)) - f(q)}{h}.$$

There are at least two big consequences of the above results. The first is that you can find charts so that the integral curves of any vanishing vector field look like one of the axes.

Theorem 47. *Suppose $X \in \mathcal{X}(M)$ is smooth and $p \in M$. If $X_p \neq 0$, then there is a chart (x, U) around p so that $X = \frac{\partial}{\partial x^1}$ in U.*

Proof. It's enough to show the theorem for $M = \mathbb{R}^n$ with standard coordinates $(t^1, \ldots, t^n)$ and $p = 0$. By rotating and scaling the coordinate system we can also assume that $X_p = \frac{\partial}{\partial t^1}|_p$.

The essential idea is that in a neighborhood of 0 there is a unique integral curve through each $(0, a^2, \ldots, a^n)$, so we'll use the time of the flow for the first coordinate and the remaining a^j for the others.

Let ϕ_t be the family of diffeomorphisms for X and consider the map

$$\psi(a^1, a^2, \ldots, a^n) = \phi_{a^1}(0, a^2, \ldots, a^n).$$

We now compute

$$\begin{aligned}\psi_*\left(\frac{\partial}{\partial t^1}|_a\right)(f) &= \frac{\partial}{\partial t^1}|_a(f\circ\psi) \\ &= \lim_{h\to 0}\frac{1}{h}\left(f(\psi(a^1+h,a^2,\dots,a^n)) - f(\psi(a))\right) \\ &= \lim_{h\to 0}\frac{1}{h}\left[f(\phi_{a^1+h}(0,a^2,\dots,a^n)) - f(\psi(a))\right] \\ &= \lim_{h\to 0}\frac{1}{h}\left[f(\phi_h(\psi(a))) - f(\psi(a))\right] = (Xf)(\psi(a)),\end{aligned}$$

and $\psi_*\left(\frac{\partial}{\partial t^j}|_0\right) = \frac{\partial}{\partial t^j}|_0$, so that $(\psi_*)_0 = I$ is nonsingular and thus ψ^{-1} is a coordinate system. ☐

It's worth asking (and later we'll answer) whether there is a similar theorem for a larger number of vector fields.

The other big consequence is the definition of the Lie derivative, which we'll get to later.

7.2. **Derivations and the Lie bracket.** We can think of a smooth vector field as giving us an operator on $C^\infty(M)$. If $X\in\mathcal{X}(M)$ and $f\in C^\infty(M)$, we define

$$Xf\in C^\infty(M),\quad (Xf)(p) = X_p(f).$$

Because the value of a derivation on f is determined by the local behavior of the function, the same is true for Xf, i.e., if $U\subset M$ is open, we have

$$(Xf)|_U = X\left(f|_U\right).$$

Smooth vector fields provide derivations on $C^\infty(M)$:

$$X(fg) = fX(g) + gX(f).$$

Two of the implications in the following proposition are proved by moving to coordinates; the other implication follows from an extension argument.

Proposition 48 (Lee, Proposition 8.14). *Let M be a smooth manifold and X a (not necessarily smooth) vector field on M. The following are equivalent:*

(a) X is smooth,
(b) For every $f\in C^\infty(M)$, the function Xf is smooth, and
(c) For every open subset $U\subset M$ and every $f\in C^\infty(U)$, Xf is smooth on U.

Lemma 49 (Lee, Proposition 8.15). *Let M be a smooth manifold with boundary. A map $D: C^\infty(M)\to C^\infty(M)$ isa derivation if and only if it is of the form $Df = Xf$ for some smooth vector field $X\in\mathcal{X}(M)$.*

Proof. Every smooth vector field yields a derivation, so we must only show the "only if" part of the lemma. We must therefore concoct X from D; its value at p must be the derivation at p whose action on any smooth function f is given by

$$X_pf = (Df)(p).$$

As D is linear, this expression depends linearly on f; as D is a derivation on $C^\infty(M)$, X_p is a derivation at p and so X is a (not necessarily smooth) vector field. That it is smooth follows from the previous lemma. ☐

Now we describe a way of combining two vector fields to obtain a new one. Let $X, Y \in \mathcal{X}(M)$. Given a smooth function $f : M \to \mathbb{R}$, we can apply X to f to get a smooth function Xf, to which we can then apply Y and obtain another smooth function $YXf = Y(Xf)$. In general, however, the operation $f \mapsto YXf$ does not in general satisfy the product rule and therefore cannot be a vector field. (Why? Second derivatives.) On the other hand, we can reverse the order of application to obtain yet another smooth function XYf. Taking their difference yields an operator $[X, Y] : C^\infty(M) \to C^\infty(M)$ called the *Lie bracket of X and Y* or the *commutator of X and Y*:

$$[X, Y]f = XYf - YXf.$$

This operator *is* a vector field.

Lemma 50 (Lee, Lemma 8.25)**.** *The Lie bracket of any pair of smooth vector fields is a smooth vector field.*

Proof. By the previous lemma, it suffices to show that, given $X, Y \in \mathcal{X}(M)$, $[X, Y]$ is a derivation on $C^\infty(M)$. We'll do this in two ways; the first is coordinate-free while the second works in a chart and emphasizes what is going on.

Let $f, g \in C^\infty(M)$, then we have

$$\begin{aligned}
[X, Y](fg) &= X\left(Y(fg)\right) - Y\left(X(fg)\right) \\
&= X\left(fYg + gYf\right) - Y\left(fXg + gXf\right) \\
&= (Xf)(Yg) + f(XYg) + (Xg)(Yf) + g(XYf) \\
&\quad - (Yf)(Xg) - f(YXg) - (Yg)(Xf) - g(YXf) \\
&= f[X, Y]g + g[X, Y]f,
\end{aligned}$$

so $[X, Y]$ is a derivation and thus a vector field.

We now fix a coordinate chart and write the vector fields in terms of a basis, i.e., $X = \sum_{i=1}^n X^i \frac{\partial}{\partial x^i}$, $Y = \sum_{j=1}^n Y^j \frac{\partial}{\partial x^j}$. Here X^i and Y^j are smooth functions in the chart. We then take $f, g \in C^\infty(M)$ and compute

$$\begin{aligned}
[X, Y](fg) &= \left[\sum_i X^i \frac{\partial}{\partial x^i}, \sum_j Y^j \frac{\partial}{\partial x^j}\right](fg) \\
&= \sum_{i,j=1}^n \left[X^i \frac{\partial}{\partial x^i}, Y^j \frac{\partial}{\partial x^j}\right](fg).
\end{aligned}$$

By linearity it therefore suffices to consider one of the terms above: For any $f \in C^\infty(M)$,

$$\begin{aligned}
\left[X^i \frac{\partial}{\partial x^i}, Y^j \frac{\partial}{\partial x^j}\right](f) &= X^i \frac{\partial}{\partial x^i}\left(Y^j \frac{\partial}{\partial x^j}\right)(f) - Y^j \frac{\partial}{\partial x^j}\left(X^i \frac{\partial}{\partial x^i}\right)(f) \\
&= X^i \left(\frac{\partial Y^j}{\partial x^i}\frac{\partial f}{\partial x^j} + Y^j \frac{\partial^2 f}{\partial x^i \partial x^j}\right) - Y^j \left(\frac{\partial X^i}{\partial x^j}\frac{\partial f}{\partial x^i} + X^i \frac{\partial^2 f}{\partial x^j \partial x^i}\right) \\
&= \left(X^i \frac{\partial Y^j}{\partial x^i}\frac{\partial}{\partial x^j} - Y^j \frac{\partial X^i}{\partial x^j}\frac{\partial}{\partial x^i}\right)(f).
\end{aligned}$$

Applying this expression to fg then shows that it is a derivation. $\square$

A consequence of the second proof given above is that if X^i are the components of X and Y^j are the components of Y in a chart, then j-th component of $[X, Y]$ in the chart is

$$\sum_{i=1}^{n} \left(X^i \frac{\partial Y^j}{\partial x^i} - Y^i \frac{\partial X^j}{\partial x^i} \right).$$

We collect a few properties of the Lie bracket:

Proposition 51. *Suppose $X, Y, Z \in \mathcal{X}(M)$.*
(a) BILINEARITY: For $a, b \in \mathbb{R}$,

$$[aX + bY, Z] = a[X, Z] + b[Y, Z],$$
$$[X, aY + bZ] = a[X, Y] + b[X, Z].$$

(b) ANTISYMMETRY: $[X, Y] = -[Y, X]$.
(c) JACOBI IDENTITY:

$$[X, [Y, Z]] + [Y, [Z, X]] + [Z, [X, Y]] = 0.$$

(d) For $f, g \in C^\infty(M)$,

$$[fX, gY] = fg[X, Y] + (fXg)Y - (gYf)X.$$

Proof. The only one of the above that is not quick exercises is the Jacobi identity. For that we compute (essentially symbolically):

$$\begin{aligned}
&[X, [Y, Z]] + [Y, [Z, X]] + [Z, [X, Y]]\, f \\
&\quad = X[Y, Z]f - [Y, Z]Xf + Y[Z, X]f - [Z, X]Yf + Z[X, Y]f - [X, Y]Zf \\
&\quad = XYZf - XZYf - YZXf + ZYXf + YZXf - YXZf \\
&\qquad - ZXYf + XZYf + ZXYf - ZYXf - XYZf + YXZf \\
&\quad = 0.
\end{aligned}$$

□

Suppose that X is a vector field and ϕ_t is the corresponding 1-parameter family of diffeomorphisms. We define the Lie derivative

Appendix A. Rewrites

Rewriting Sections 4 and 5 to further emphasize the cotangent bundle.

A.1. The cotangent bundle.

A.1.1. *Some words about vector bundles.* Suppose B is an n-dimensional manifold and E is an $(n + k)$-dimensional manifold and there is a (continuous) map $\pi : E \to B$. We define the notion of a vector bundle, which is where we think of each point in a manifold B (which stands for "base") as having a finite dimensional vector space attached to it. Even though we'll demand that all of these vector spaces be abstractly isomorphic (i.e., they'll have the same dimension), you should think of them as *distinct* vector spaces as any isomorphism you come up with typically depends on a lot of choices.

Definition 52. We say E is a vector bundle over B if
(1) π is surjective,
(2) $\pi^{-1}(p)$ is a k-dimensional vector space for each $p \in B$, and

(3) for each $p \in B$, there is chart (x, U) around p in B and a diffeomorphism $\varphi : \pi^{-1}(U) \to x(U) \times \mathbb{R}^n$ that restricts to be a vector space isomorphism on each fiber, i.e., $\varphi : \pi^{-1}(p) \to \{x(p)\} \times \mathbb{R}^n$ is a a vector space isomorphism for each $p \in U$.

In the above, you think of the space $\{x(p)\} \times \mathbb{R}^n$ as a vector space equipped with the addition and scaling laws of

$$\begin{aligned} &(x(p), v) + (x(p), w) = (x(p), v + w), \\ &c(x(p), v) = (x(p), cv). \end{aligned}$$

Definition 53. A *section* of a vector bundle E over B is a (smooth, continuous, etc. depending on category) map $s : B \to E$ so that $\pi \circ s = \mathrm{Id}_B$. In other words, it is a map $s : B \to E$ so that for each $p \in B$, $s(p) \in \pi^{-1}(p)$.

Definition 54. A bundle map of two vector bundles is one that preserves the fibers and is linear on each fiber.

You can think of a bundle map as secretly being two maps. In other words, if E_1 is a vector bundle over B_1, E_2 is a vector bundle over B_2, and $f : E_1 \to E_2$ is a bundle map, it induces a map $f_B : B_1 \to B_2$ making the following diagram commute:

$$\begin{tikzcd} E_1 \arrow[r, "f"] \arrow[d, "\pi_1"] & E_2 \arrow[d, "\pi_2"] \\ B_1 \arrow[r, "f_B"] & B_2 \end{tikzcd}$$

A.1.2. *The cotangent space at a point in* $\mathbb{R}^n$. For an open set $U \subset \mathbb{R}^n$, we'd like to define T_p^*U to be the vector space $\{p\} \times (\mathbb{R}^n)^*$, where $(\mathbb{R}^n)^*$ denotes the space of row vectors (with n entries) and the operations are given by

$$(p, \xi) + (p, \eta) = (p, \eta + \xi), \quad c.(p, \xi) = (p, c\xi).$$

(In other words, the p is there only to remind you not to try to add covectors corresponding to different points.)

Instead, we'll take a different approach and define T_p^*U more intrinsically and then show it is isomorphic to the space above. For $p \in U$, we define the ideal of smooth functions vanishing at p:

$$\mathcal{I}_p = \{f : U \to \mathbb{R} \mid f \text{ is smooth and } f(p) = 0\}.$$

We let $\mathcal{I}_p^2$ denote the square of this ideal, i.e.,

$$\mathcal{I}_p^2 = \left\{ \sum_{j=1}^k f_j g_j \mid f_j, g_j \in \mathcal{I}_p \right\}.$$

We now define the cotangent space to U at p by

$$T_p^*U = \mathcal{I}_p / \mathcal{I}_p^2.$$

Proposition 55. *If $U \subset \mathbb{R}^n$ is open, then T_p^*U is an n-dimensional vector space.*

The proposition follows from two lemmas, which together show that Taylor's theorem provides a basis for T_p^*M.

Lemma 56. *If $f \in \mathcal{I}_p$, then there is a neighborhood of p and a constant C so that $|f(x)| \leq C|x-p|$ for all x in that neighborhood. In particular, if $f \in \mathcal{I}_p^2$, then $|f(x)| \leq C|x-p|^2$ in a neighborhood of p.*

Proof. For $f \in \mathcal{I}_p$, Taylor's theorem provides that

$$f(x) = \sum_{j=1}^{n} \partial_j f(p)(x^j - p^j) + R_2(x).$$

As $x^j - p^j \in \mathcal{I}_p$, the naïve bounds for Taylor's remainder formula and the first two terms here prove the result. The result for $\mathcal{I}_p^2$ follows from the estimate for $\mathcal{I}_p$. $\square$

The following lemma provides uniqueness of the first Taylor polynomial (though it adapts easily to higher Taylor polynomials).

Lemma 57. *If $U \subset \mathbb{R}^n$ is open, $a \in U$ and $f : U \to \mathbb{R}$ is smooth, and if $f(x) = P(x) + R(x)$, where $P(x)$ is a polynomial of degree 1 and $|R(x)| \leq C|x-a|^2$ for x near a, then $P(x)$ is the first-order Taylor polynomial of f at a.*

Proof. Writing $f(x) = P_1(x) + R_1(x)$ as in Taylor's theorem, we know that $R_1(x)$ satisfies the same bound. We now consider

$$|P(x) - P_1(x)| \leq |R_1(x)| + |R(x)| \leq C|x-a|^2.$$

As $P(x) - P_1(x)$ is a first degree polynomial vanishing quadratically at a, we must have that it vanishes identically, i.e., $P(x) = P_1(x)$, as claimed. $\square$

Proof of Proposition 55. That T_p^*U is a vector space is clear because $\mathcal{I}_p$ is a vector space and $\mathcal{I}_p^2$ is a subspace. We must only show it has dimension n.

If $(p^1, \dots, p^n)$ denote the coordinates of p and $(x^1, \dots, x^n)$ are the coordinate functions on $\mathbb{R}^n$, we claim that $x^1 - p^1, \dots, x^n - p^n$ form a basis for T_p^*U. That they span $\mathcal{I}_p/\mathcal{I}_p^2$ follows from Taylor's theorem; their linear independence follows from the uniqueness statement in Lemma 57. $\square$

From the proof of Proposition 55, we see that the coordinate functions $x^1 - p^1, \dots, x^n - p^n$ provide a basis for $\mathcal{I}_p/\mathcal{I}_p^2$. We introduce special notation for this basis; we let dx^j denote the image of $x^j - p^j$ in T_p^*U. Indeed, given any smooth function f, we can consider its image $df_p = f - f(p)$ in T_p^*U. In terms of the basis above, Taylor's theorem tells us that

$$df_p = \sum_{j=1}^{n} \frac{\partial f}{\partial x^j}(p) dx^j.$$

Now suppose that $F : U \to V$ is a smooth map, where $U \subset \mathbb{R}^n$ and $V \subset \mathbb{R}^m$ are open. We know that smooth functions pull back to smooth functions, so if $f \in C^\infty(V)$, we have $F^*f = f \circ F \in C^\infty(U)$.