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.) $\square$

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

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

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. $\square$

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

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

$\square$

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 $(x^1(p), x^2(p), \dots, x^n(p))$ 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 $j = 1, \dots, n+1$:

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

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.

$$\begin{tikzcd} M \arrow[r, "f"] & N \arrow[d, "y"] \\ x(U) \arrow[u, "x^{-1}"] \arrow[r] & y(V) \end{tikzcd}$$

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) $\displaystyle\sum_{\alpha\in A} \psi_\alpha(x) = 1$ 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 $\frac{\partial}{\partial x^j}$, $j = 1, \ldots, n$ 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. □

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. $\square$

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. $\square$

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:

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

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 $T_p M$ 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

$$T(\alpha v_1 + \beta v_2, w) = \alpha T(v_1, w) + \beta T(v_2, w),$$
$$T(v, \alpha w_1 + \beta w_2) = \alpha T(v, w_1) + \beta T(v, w_2).$$

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

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

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 $\varphi(v, w) = v \otimes w$.

[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*

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

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

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

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

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

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

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

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

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

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

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

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 $f(x_\infty) = x_\infty$.

For uniqueness, note that if $\tilde{x}_\infty$ is another fixed point, we have

$$d(x_\infty, \tilde{x}_\infty) = d(f(x_\infty), f(\tilde{x}_\infty)) \leq Cd(x_\infty, \tilde{x}_\infty).$$

As $C < 1$, we must have $d(x_\infty, \tilde{x}_\infty) = 0$, i.e., $x_\infty = \tilde{x}_\infty$. ☐

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.

[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 $y_1, y_2 \in U$.

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
(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

$$\gamma_x'(t) = f(t, \gamma_x(t)),$$
$$\gamma_x(t_0) = x.$$

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 $\gamma : (t_0 - b, t_0 + b) \times U \to \mathbb{R}^n$.*

Well, ok, we'll prove it in the notes but not in class. See Appendix B.

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

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

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. Suppose that X is a vector field and ϕ_t is the corresponding 1-parameter family of diffeomorphisms. We define the Lie derivative of a vector field $Y \in \mathcal{X}(M)$ by

$$(L_X Y)_p = \lim_{h\to 0}\frac{1}{h}\left[Y_p - \left((\phi_h)_* Y\right)_p\right].$$

Note that $\left((\phi_h)_* Y\right)_p$ means

$$(\phi_h)_*\left(Y_{\phi_h^{-1}(p)}\right) = (\phi_h)_*(Y_{\phi_{-h}(p)}).$$

While we're at it, let's also define the Lie derivative of a 1-form.[8] For a 1-form ω and a smooth vector field X, we define

$$(L_X\omega)_p = \lim_{h\to 0}\frac{1}{h}\left[((\phi_h^*\omega)_p - \omega_p\right].$$

We also define the Lie derivative of functions by

$$(L_X f)(p) = X(f)(p) = \lim_{h\to 0}\frac{f(\phi_h(p)) - f(p)}{h}.$$

Proposition 48. *Suppose $X, Y, Z \in \mathcal{X}(M)$, $f \in C^\infty(M)$, and $\omega, \eta \in \Omega^1(M)$.*

(1) $L_X(Y + Z) = L_X Y + L_X Z$,
(2) $L_X(\omega + \eta) = L_X\omega + L_X\eta$,
(3) $L_X(fY) = (Xf)Y + fL_X Y$,
(4) $L_X(f\omega) = (Xf)\omega + fL_X\omega$, *and*
(5) $X(\omega(Y)) = L_X(\omega(Y)) = (L_X\omega)(Y) + \omega(L_X Y)$.

[8]We haven't defined it in this text yet, but a 1-form is a smooth section of the *cotangent* bundle. In other words, a 1-form ω is a smooth map $\omega : M \to T^*M$ so that $\pi \circ \omega = \mathrm{Id}_M$. The space of 1-forms on M is denoted $\Omega^1(M)$.

Proof. The first two are easy. All of the remaining use the same trick. Let's prove the last one.

$$\begin{aligned}
L_X(\omega(Y))(p) &= \lim_{h\to 0}\frac{1}{h}\left[\omega(Y)(\phi_h(p)) - \omega(Y)(p)\right]\\
&= \lim_{h\to 0}\frac{1}{h}\left[\phi_h^*(\omega(Y))(p) - \omega(Y)(p)\right]\\
&= \lim_{h\to 0}\frac{1}{h}\left(\left[\omega_{\phi_h(p)}(Y_{\phi_h(p)}) - \omega_{\phi_h(p)}((\phi_h)_*Y_p)\right] + \left[\omega_{\phi_h(p)}((\phi_h)_*Y_p) - \omega_p(Y_p)\right]\right)\\
&= \lim_{h\to 0}\left(\omega_{\phi_h(p)}\left(\frac{Y_{\phi_h(p)} - (\phi_h)_*(Y_p)}{h}\right) + \left(\frac{(\phi_h^*\omega - \omega)_p}{h}(Y_p)\right)\right)\\
&= \omega(L_xY)(p) + (L_X\omega)(Y)(p).
\end{aligned}$$

$\square$

We now compute in coordinates. It's easiest to do it for 1-forms and functions first, then deduce the formula for vector fields. Recall that in coordinates,

$$f^*(dy^i) = \sum_j \frac{\partial(y^i\circ f)}{\partial x^j}\,dx^j.$$

In particular, in a coordinate system (x, U) and $X = \sum_i a^i \frac{\partial}{\partial x^i}$, we have

$$\phi_h^*(dx^i) = \lim_{h\to 0}\frac{1}{h}\left[\sum_j \frac{\partial(x^i\circ\phi_h)}{\partial x^j}\,dx^j - dx^i\right].$$

The coefficient of dx^j is therefore

$$\begin{aligned}
\lim_{h\to 0}\frac{1}{h}\left[\frac{\partial(x^i\circ\phi_h)}{\partial x^j} - \delta^i_j\right] &= \lim_{h\to 0}\frac{1}{h}\left[\frac{\partial(x^i\circ\phi_h)}{\partial x^j} - \frac{\partial x^i}{\partial x^j}\right]\\
&= \frac{\partial}{\partial x^j}|_p \lim_{h\to 0}\frac{1}{h}\left[x^i\circ\phi_h - x^i\circ\phi_0\right] = \frac{\partial a^i}{\partial x^j}.
\end{aligned}$$

We can thus conclude that, for $X = \sum a^i \frac{\partial}{\partial x^i}$,

$$L_X dx^i = \sum_j \frac{\partial a^i}{\partial x^j}\,dx^j.$$

Thus if $\omega = \sum \omega_i dx^i$, we have

$$L_X\omega = \sum_i \left[X(\omega_i)dx^i + \omega_i L_X dx^i\right],$$

from which one can derive a longer, more detailed formula.

To compute $L_X Y$, we first find $L_X\left(\frac{\partial}{\partial x^j}\right)$. We have, writing $X = \sum_i a^i \frac{\partial}{\partial x^i}$,

$$\begin{aligned} 0 = L_X\left(dx^i(\frac{\partial}{\partial x^j})\right) &= L_X(dx^i)(\frac{\partial}{\partial x^j}) + dx^i(L_X \frac{\partial}{\partial x^j}) \\ &= \sum_k \frac{\partial a^i}{\partial x^k} dx^k(\frac{\partial}{\partial x^j}) + dx^i(L_X \frac{\partial}{\partial x^j}) \\ &= \frac{\partial a^i}{\partial x^j} + dx^i(L_X \frac{\partial}{\partial x^j}), \end{aligned}$$

so the $\partial/\partial x^i$ component is $-\partial a^i/\partial x^j$. In other words,

$$L_X \frac{\partial}{\partial x^j} = -\sum_i \frac{\partial a^i}{\partial x^j} \frac{\partial}{\partial x^i}.$$

Now, if $Y = \sum_i b^i \frac{\partial}{\partial x^i}$, we have

$$\begin{aligned} L_X Y &= \sum_i X(b^i) \frac{\partial}{\partial x^i} + \sum_i b^i L_X \frac{\partial}{\partial x^i} \\ &= \sum_{i,j} \left(a^j \frac{\partial b^i}{\partial x^j} \frac{\partial}{\partial x^i} - b^i \frac{\partial a^j}{\partial x^i} \frac{\partial}{\partial x^j} \right) \\ &= \sum_{i,j=1}^{n} \left(a^i \frac{\partial b^j}{\partial x^i} - b^i \frac{\partial a^j}{\partial x^i} \right) \frac{\partial}{\partial x^j}. \end{aligned}$$

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 49 (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 50 (Lee, Proposition 8.15)**.** *Let M be a smooth manifold with boundary. A map $D : C^\infty(M) \to C^\infty(M)$ is a 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

$$X_p f = (Df)(p).$$

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 51 (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. □

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

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

In particular, the Lie bracket $[X, Y]$ is also the Lie derivative $L_X Y$.

We collect a few properties of the Lie bracket:

Proposition 52. *Suppose $X, Y, Z \in \mathcal{X}(M)$.*

(a) BILINEARITY: For $a, b \in \mathbb{R}$,

$$\begin{aligned}[aX + bY, Z] &= a[X, Z] + b[Y, Z], \\ [X, aY + bZ] &= a[X, Y] + b[X, Z].\end{aligned}$$

(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):

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

□

Because $L_X Y = [X, Y]$, we now have the following non-obvious results about the Lie derivative:

Corollary 53. *(1) $L_X X = 0$,*

(2) $L_X Y = -L_Y X$,

(3) $L_X [Y, Z] = [Y, L_X Z] + [L_X Y, Z]$.

7.3. **A geometric interpretation of the Lie bracket.** In what follows we specialize to the case where $X, Y \in \mathcal{X}(M)$ have compact support in order to guarantee that the flows exist for all time.

Suppose $X \in \mathcal{X}(M)$ generates the family ϕ_t of diffeomorphisms while Y generates ψ_s. The aim of this section is to show that $[X, Y] = 0$ if and only if $\phi_t \circ \psi_s \circ \phi_{-t} \circ \psi_{-s} = \mathrm{Id}_M$ for all t, s.

The next lemma has an important consequence: if $\alpha : M \to M$ is a diffeomorphism, then $\alpha_* X = X$ if and only if $\phi_t \circ \alpha = \alpha \circ \phi_t$. In particular, if ϕ_t is the flow generated by X and ψ_t by Y, then $(\phi_t)_* Y = Y$ if and only if $\phi_t \circ \psi_s = \psi_s \circ \phi_t$.

Lemma 54. *If $\alpha : M \to N$ is a diffeomorphism and $X \in \mathcal{X}(M)$ generates the family ϕ_t, then the vector field $\alpha_* X$ generates $\alpha \circ \phi_t \circ \alpha^{-1}$.*

(Recall that $(\alpha_* X)_q = (D\alpha)_{\alpha^{-1}(q)} X_{\alpha^{-1}(q)}$.)

Proof. This follows essentially immediately from the uniqueness part of Picard iteration. Indeed, we let ψ_t denote the flow generated by $\alpha_* X$, so that ψ_t satisfies

$$\begin{aligned} \frac{d}{dt}\psi_t(q) &= (\alpha_* X)_{\psi_t(q)} = (D_\alpha)_{\alpha^{-1}(\psi_t(q))} X_{\alpha^{-1}(\psi_t(q))}, \\ \psi_0(q) &= q. \end{aligned}$$

Similarly, we now consider the flow $\alpha \circ \phi_t \circ \alpha^{-1}$, which (regarded as a map $\mathbb{R} \to N$) satisfies

$$\begin{aligned} \frac{d}{dt}(\alpha \circ \phi_t \circ \alpha^{-1})(q) &= (D\alpha)_{\phi_t(\alpha^{-1}(q))} \frac{d}{dt}\phi_t(\alpha^{-1}(q)) = (D\alpha)_{\phi_t(\alpha^{-1}(q))} X_{\phi_t(\alpha^{-1}(q))} \\ &= (D\alpha)_{\alpha^{-1}((\alpha\circ\phi_t\circ\alpha^{-1})(q))} X_{\alpha^{-1}((\alpha\circ\phi_t\circ\alpha^{-1})(q))}, \\ \alpha \circ \phi_0 \circ \alpha^{-1}(q) &= q. \end{aligned}$$

so that the two flows satisfy the same differential equation with the same initial condition and thus must agree. $\square$

Theorem 55. *If $X, Y \in \mathcal{X}(M)$ are compactly supported (or otherwise generate globally defined flows), X generates ϕ_t, and Y generates ψ_s, then $[X, Y] = 0$ if and only if $\phi_t \circ \psi_s = \psi_s \circ \phi_t$ for all t, s.*

Proof. If $\phi_t \circ \psi_s = \psi_s \circ \phi_t$ then $(\phi_t)_* Y = Y$, so $L_X Y = [X, Y] = 0$.

For the other direction, suppose that $L_X Y = [X, Y] = 0$, so for each $p \in M$,

$$\lim_{h \to 0} \frac{Y_p - ((\phi_h)_* Y)_p}{h} = 0.$$

Now, given $p \in M$, we define the curve $c : \mathbb{R} \to T_p M$ by

$$c(t) = ((\phi_t)_* Y)_p \, .$$

We claim that c is constant, so that $(\phi_t)_* Y = Y$, which is equivalent to the desired claim. We compute (using that the tangent space to $T_p M$ can be canonically identified with $T_p M$

itself):

$$\begin{aligned} c'(t) &= \lim_{h\to 0} \frac{1}{h}\left[((\phi_{t+h})_* Y)_p - ((\phi_t)_* Y)_p \right] \\ &= \lim_{h\to 0} \frac{1}{h}\left[(D\phi_{t+h})_{\phi_{-t-h}(p)} Y_{\phi_{-t-h}(p)} - (D\phi_t)_{\phi_{-t}(p)} Y_{\phi_{-t}(p)} \right] \\ &= \lim_{h\to 0} (D\phi_t)_{\phi_{-t}(p)} \left[\frac{(D\phi_h)_{\phi_{-t-h}(p)} Y_{\phi_{-t-h}(p)} - Y_{\phi_{-t}(p)}}{h} \right] \\ &= -(D\phi_t)_{\phi_{-t}(p)} (L_X Y)_{\phi_{-t}(p)} = 0. \end{aligned}$$

□

8. Integral submanifolds and the Frobenius theorem

Now, given two vector fields X_1 and X_2, must they look like $\frac{\partial}{\partial x^1}$ and $\frac{\partial}{\partial x^2}$? The answer is definitely no, for an obvious reason: coordinate vector fields commute (because second partial derivatives commute), but we needn't have $[X_1, X_2] = 0$.

More interesting is that this condition turns out to be sufficient.

Proposition 56. *If $X_i, i = 1, \dots, k$ are smooth vector fields that are linearly independent at p, then $[X_i, X_j] = 0$ for all i and j if and only if there is a neighborhood U of p and a smooth map $\psi : U \to \mathbb{R}^n$ so that $\psi_*(X_i) = \frac{\partial}{\partial x^i}$.*

Proof. The idea is to take X_1, transform so that X_1 looks locally like $\frac{\partial}{\partial x^1}$, then arrange X_2 so that it looks "perpendicular" at p, then try to push the transversal forward with ψ_t to fix X_2. Did we screw up x_1? Not if the flows commute, so not if the vector fields commute.

By working in a chart, we can assume $M = \mathbb{R}^n$ and $p = 0$. As the X_i are linearly independent, we can use a linear change of coordinates to ensure that $(X_i)_0 = \frac{\partial}{\partial x^i}|_0$.

Let ϕ^i_t denote the flow generated by X_i and, for $(t^1, \dots, t^n)$ sufficiently close to 0, define

$$\chi(t^1, \dots, t^n) = \phi^1_{t^1}\left(\phi^2_{t^2}\left(\dots\left(\phi^k_{t^k}\left(0, \dots, 0, t^{k+1}, \dots, t^n\right)\right)\right)\right).$$

Observe that taking partial derivatives yields

$$D\chi_0\left(\frac{\partial}{\partial x^j}|_0\right) = \begin{cases} (X_i)_0 = \frac{\partial}{\partial x^i}|_0 & i = 1, \dots, k \\ \frac{\partial}{\partial x^i}|_0 & i > k \end{cases}$$

so that χ^{-1} is a coordinate system in a neighborhood of 0.

In this coordinate system, our construction implies that $X_1 = \frac{\partial}{\partial x^1}$. Now, since $[X_i, X_j] = 0$ for $i, j \leq k$, the flows ϕ^i_t and ϕ^j_t commute, so in fact we can rewrite the formula for χ to ensure that $\phi^j_{t^j}$ is the outermost flow and hence in this coordinate system we have $X_j = \frac{\partial}{\partial x^j}$. □

8.1. Submanifolds and distributions.

Definition 57. We say that $F : M \to N$ is an *immersion* if DF_p is injective for all $p \in M$. A map $F : M \to N$ is a *submersion* if DF_p is surjective for all $p \in M$. A map $F : M \to N$ is an *embedding* if it is an immersion and a homeomorphism onto its image.

We won't prove this in class, but it's useful to know:

Theorem 58 (Whitney embedding theorem)**.** *Every C^∞ n-dimensional smooth manifold can be smoothly embedded into $\mathbb{R}^{2n}$.*

We probably won't do this in class, but here is a weaker related theorem:

Theorem 59. *Every compact smooth manifold can be embedded into $\mathbb{R}^N$ for sufficiently large N.*

Proof. The manifold M is compact, so we can cover it by a finite number of charts $(x_1, U_1), \ldots, (x_k, U_k)$. We choose U_i' open sets so that $\overline{U_i'} \subset U_i$ and so that the U_i' still cover M. Find $\psi_i : M \to [0, 1]$ so that $\psi_i \equiv 1$ on U_i' and $\operatorname{supp} \psi_i \subset U_i$.

Let $F : M \to \mathbb{R}^{nk+k}$ be defined by

$$F(p) = (\psi_1(p)x_1(p), \psi_2(p)x_2(p), \ldots, \psi_k(p)x_k(p), \psi_1(p), \ldots, \psi_k(p)) .$$

We claim that F is an embedding. We first show it is an immersion.

Indeed, any $p \in M$ lies in U_i' for some i and so the Jacobian of F contains as a submatrix $\frac{\partial x_i^\alpha}{\partial x_i^\beta} = \mathrm{Id}_{n\times n}$ and so has rank n.

To show that F is injective is a straightforward exercise; that it is a homeomorphism onto its image follows because F is an injective continuous function from a compact set to a Hausdorff space. □

We now return to the original problem.

Definition 60. A k-dimensional distribution Δ on M is a k-dimensional sub-bundle of TM.

In other words, a k-dimensional distribution Δ on M is an assignment of a k-dimensional subspace of T_pM to each $p \in M$ in such a way that the subspaces fit together smoothly. One way of thinking about this is we demand that Δ is a k-dimensional vector bundle over M equipped with an embedding $\Delta \hookrightarrow TM$ commuting with the projections:

$$\begin{tikzcd} \Delta \arrow[r, hook, "\iota"] \arrow[d, "\pi"] & TM \arrow[d, "\pi"] \\ M \arrow[r, "\mathrm{Id}"] & M \end{tikzcd}$$

You can think of a one-dimensional distribution as generalizing the notion of a non-vanishing vector field (where X is replaced by $\operatorname{span} X$). Not all one-dimensional distributions are of this form, though (recall your construction of the Möbuis strip).

Definition 61. Given a k-dimensional distribution $\Delta \subset TM$, an integral submanifold of Δ is a map $f : N \to M$ so that f_* is injective and $f_*(TN) \subset \Delta$. (By default, the dimension of N should be the same as the dimension of Δ.)

8.2. **The Frobenius integrability theorem, take 1.** Our aim is to answer the question of when a distribution has an integral submanifold (thereby generalizing the discussion at the beginning of this section). Recall that a necessary condition is that it be closed under the Lie bracket.

Theorem 62 (Frobenius theorem)**.** *If sections of Δ are closed under the Lie bracket, then for each $p \in M$ there is a neighborhood U of p and a map $\psi : U \to \mathbb{R}^n$ so that locally Δ is the tangent space to the fibers of the projection $\mathbb{R}^n \to \mathbb{R}^{n-k}$, i.e., each fiber $\pi^{-1}(a)$ for $a \in \mathbb{R}^{n-k}$ gives an integral submanifold.*

Lemma 63. *Suppose $F : M \to N$ is an immersion. Any vector field Y on N that lies in the image of DF along F is F-related to a vector field on M.*

(Recall that for $F : M \to N$, $X \in \mathcal{X}(M)$ is F-related to $Y \in \mathcal{X}(N)$ if $DF_p(X_p) = Y_{F(p)}$ for all $p \in M$.)

Proof. For $p \in M$, as DF_p is injective, we let X_p denote the unique element of T_pM so that $DF_p(X_p) = Y_{F(p)}$. The vector field X is smooth because F is an immersion (i.e., because you can choose coordinates so that $y \circ x^{-1}(a^1, \dots a^n) = (a^1, \dots, a^n, 0, \dots, 0)$). □

Proposition 64. *Suppose X_i and Y_i are F-related for $i = 1, 2$. Then $[X_1, X_2]$ is F-related to $[Y_1, Y_2]$.*

We'll prove this proposition in a bit. First let's prove the theorem assuming the proposition.

Proof of theorem. We can assume $M = \mathbb{R}^n$ and $p = 0$, and $\Delta_0 = \operatorname{span}\left(\frac{\partial}{\partial x^1}, \dots, \frac{\partial}{\partial x^k}\right)$. Let $\pi_L : \mathbb{R}^n \to \mathbb{R}^k$ denote the projection onto the first k factors, so that $(D\pi_L)_0 : \Delta_0 \to \mathbb{R}^k$ is an isomorphism and $D\pi_L : \Delta \to \mathbb{R}^k$ is also one-to-one in a neighborhood of 0.

For each q in this neighborhood, we can then choose unique $(X_1)_q, \dots, (X_k)_q \in \Delta_q$ so that

$$(D\pi_L)_q(X_j)_q = \frac{\partial}{\partial x^j}|_{\pi_L(q)},$$

for $j = 1, \dots, k$. In particular, X_j and $\frac{\partial}{\partial x^j}$ are π_L-related and thus by the proposition,

$$(\pi_L)_*[X_i, X_j] = \left[\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right] = 0,$$

so $[X_i, X_j] = 0$ as $[X_i, X_j] \in \Delta_q$ and $D\pi_L$ is injective on Δ_q.

By the proposition at the beginning of this section, X_i are locally coordinate vector fields, i.e., there is some $\psi : U \to \mathbb{R}^n$ with the desired property. □

We now prove the remaining proposition.

Proof of proposition. First observe that if $g : N \to \mathbb{R}$ is smooth and X and Y are F-related, then

$$\begin{aligned} ((Yg) \circ F)(p) &= (dg)_{F(p)}(Y_{F(p)}) \\ &= (dg)_{F(p)}(DF_p X_p) = d(g \circ F)_p(X_p) = X_p(g \circ F). \end{aligned}$$

We apply this identity repeatedly and consider $([Y_1, Y_2]g) \circ F$:

$$\begin{aligned} ([Y_1, Y_2]g) \circ F &= (Y_1(Y_2 g)) \circ F - (Y_2(Y_1 g)) \circ F \\ &= X_1((Y_2 g) \circ F) - X_2((Y_1 g) \circ F) \\ &= X_1(X_2(g \circ F)) - X_2(X_1(g \circ F)) \\ &= [X_1, X_2](g \circ F) = F_*([X_1, X_2])(g), \end{aligned}$$

as desired. □

Note, there's generally no global map; pursuing this direction leads to the theory of foliations. (Consider a line with irrational slope on the torus.)

As an example, consider when the distribution

$$\Delta = \operatorname{span}\left(\frac{\partial}{\partial x} + f(x, y, z)\frac{\partial}{\partial z}, \frac{\partial}{\partial y} + g(x, y, z)\frac{\partial}{\partial z}\right)$$

on $\mathbb{R}^3$ integrable? Let X and Y denote these two vector fields generating Δ. We compute

$$\begin{aligned}[X,Y] &= \left(\frac{\partial g}{\partial x} + f\frac{\partial g}{\partial z}\right)\frac{\partial}{\partial z} - \left(\frac{\partial f}{\partial y} + g\frac{\partial f}{\partial g}\right)\frac{\partial}{\partial z} \\ &= \left(\frac{\partial g}{\partial x} - \frac{\partial f}{\partial y} + f\frac{\partial g}{\partial z} - g\frac{\partial f}{\partial z}\right)\frac{\partial}{\partial z}.\end{aligned}$$

We won't consider this in full detail, but observe that if f and g are independent of z, then the integrability condition for Δ agrees with the familiar vector field integrability condition

$$\frac{\partial g}{\partial x} = \frac{\partial f}{\partial y}.$$

9. Preamble to integration

There are a number of loose ends that we should clear up before we introduce differential forms and integration. In particular, we should discuss orientations, manifolds with boundary, and alternating tensor spaces.

9.1. **Orientations.** Suppose V is a finite dimensional vector space. If $L : V \to V$ is a non-singular (i.e., invertible) linear transformation, it must have strictly positive or strictly negative determinant. If the former, we say that L is "orientation-preserving," while in the latter case it is "orientation-reversing."

We now consider two *ordered* bases for V, say $\mathcal{B} = (v_1, \ldots, v_n)$ and $\mathcal{B}' = (v_1', \ldots, v_n')$ and let L be the linear transformation satisfying $Lv_k = v_k'$. We say that the two bases $\mathcal{B}$ and $\mathcal{B}'$ define the same orientation of V if $\det L > 0$ and define opposite orientations of V if $\det L < 0$.

It is straightforward to check that the property of defining the same orientation defines an equivalence relation on the set of ordered bases of V; we say that an orientation of V is a choice of equivalence class of ordered bases. (Straighforward EXERCISE: There are two equivalence classes.)

If $(V, \mu = [(v_1, \ldots, v_n)])$ and $(W, \nu = [(w_1, \ldots, w_n)])$ are two oriented n-dimensional vectors spaces and $f : V \to W$ is a linear transformation, we say that f is *orientation-preserving* if

$$[(f(v_1), \ldots, f(v_n))] = \nu \quad \text{when} \quad [(v_1, \ldots, v_n)] = \mu.$$

The standard $\mathbb{R}^n$ is equipped with a "standard orientation" $[(e_1, \ldots, e_n)]$; for a trivial bundle $X \times \mathbb{R}^n$ we can equip each fiber $\{x\} \times \mathbb{R}^n$ with the standard orientation.

Now suppose $\pi : E \to B$ is a (possibly non-trivial) k-dimensional vector bundle. An orientation μ of E is a collection of orientations μ_p on each fiber $\pi^{-1}(p)$ so that, for any connected open set $U \subset B$, if $t : \pi^{-1}(U) \to U \times \mathbb{R}^k$ is a trivialization, then t is either orientation-preserving or orientation-reversing on all of the fibers over U (here $U \times \mathbb{R}^k$ is equipped with the standard orientation). A vector bundle is orientable if it can be equipped with an orientation.

As an exercise, consider the Möbius strip. It is a 1-dimensional vector bundle over S^1 and has no non-zero sections; it cannot be orientable because in one-dimension, an orientation corresponds to an equivalence class of non-vanishing vector fields.

We say that a manifold is orientable if the vector bundle $TM \to M$ is orientable. You should check that this is equivalent to the existence of an atlas for M in which, in addition to being C^∞-related, all of the transition functions can be chosen so that the determinant of the Jacobian is positive.

9.2. **Manifolds with boundary.** At this stage we probably should have a reckoning with manifolds with boundary, but we won't. Rather than having as local models open sets in $\mathbb{R}^n$, they are modeled on open sets in $\mathbb{H}^n = [0, \infty) \times \mathbb{R}^{n-1}$. They have a distinguished subset called the boundary: for $\mathbb{H}^n$, this is $\partial \mathbb{H}^n = \{0\} \times \mathbb{R}^{n-1}$. If M is a manifold with boundary, there is a subset called ∂M defined by the property that $p \in \partial M$ if and only if for all charts around p, the image of p lies in the boundary $\partial \mathbb{H}^n$ of $\mathbb{H}^n$. (Exercise: this is well-defined.)

Many of our foundational results and definitions require no changes or only slight changes on manifolds with boundary. You are encouraged to go through and identify what changes. (Example: what additional conditions do you need in the inverse function theorem to ensure that a map $F : \mathbb{H}^n \to \mathbb{H}^n$ from a neighborhood of a point in $\partial \mathbb{H}^n$ to a neighborhood of a point in $\partial \mathbb{H}^n$ preserves the boundary?)

The definitions of tangent spaces (and related constructions) are unchanged; one can check that the tangent space at a point in the boundary of an n-dimensional manifold with boundary still has dimension n. For each point $p \in \partial M$, there is a natural inclusion map $T_p(\partial M) \hookrightarrow T_p M$ given by the derivative of the inclusion map $i : \partial M \hookrightarrow M$. A vector is tangent to the boundary if it is in the image of this inclusion map.

For the global existence one-parameter families of diffeomorphisms generated by a vector field X, one also demands that X be tangent to the boundary. (In other words, over all points $p \in \partial M$, X_p lies in the image of the inclusion $T_p(\partial M) \hookrightarrow T_p M$.) This condition ensures that the one-parameter family of diffeomorphisms preserves the boundary. If in addition X vanishes at the boundary, then the one-parameter family of diffeomorphisms fixes the boundary (i.e., $\phi_t(p) = p$ for all t and for all $p \in \partial M$).

9.3. **Alternating tensors.** We now turn our attention to another piece of linear algebra. Suppose V is an n-dimensional vector space. Recall that $T^{(0,k)}(V) = (V^*)^{\otimes k}$ is the space of $(0, k)$-tensors on V and can be identified with the space of multilinear maps from $V \times \cdots \times V$ to $\mathbb{R}$. We use this identification in our definitions below.

Definition 65. A tensor $T \in T^{(0,k)}(V)$ is *alternating* if

$$T(v_1, \dots, v_i, \dots, v_j, \dots, v_k) = 0$$

whenever $v_i = v_j$ for some $i \neq j$.

Note that if T is alternating then it is also skew-symmetric. (Let's just do this for bilinear forms here.)

$$\begin{aligned} 0 &= T(v_1 + v_2, v_1 + v_2) \\ &= T(v_1, v_1) + T(v_2, v_2) + T(v_1, v_2) + T(v_2, v_1) \\ &= T(v_1, v_2) + T(v_2, v_1). \end{aligned}$$

Note that because we are working over $\mathbb{R}$, skew-symmetric forms are also alternating. If we were working over a field of characteristic 2 (which, thankfully, we are not!), skew-symmetry is the same as symmetry and there are skew-symmetric tensors that are not alternating.

Definition 66. For an n-dimensional vector space V, $\Lambda^k(V)$ consists of all alternating $T \in T^{(0,k)}(V)$.

What is its dimension? For this it is helpful to construct the map $\mathrm{Alt} : T^{(0,k)}(V) \to \Lambda^k(V)$. Let S_n denote the symmetric group on n letters; we define an action of S_k on $T^{(0,k)}(V)$: for

$T \in T^{(0,k)}(V)$, and $\sigma \in S_k$, we define

$$(\sigma.T)(v_1, \dots, v_n) = T(v_{\sigma(1)}, v_{\sigma(2)}, \dots, v_{\sigma(n)}).$$

This action characterizes the alternating tensors:

Lemma 67. *Suppose $T \in T^{(0,k)}(V)$. T is alternating if and only if $(\sigma.T) = (\operatorname{sgn}\sigma)T$ for all $\sigma \in S_k$. (Here $\operatorname{sgn}\sigma$ is the sign of the permutation σ – it is 1 if σ can be written as a product of an even number of transpositions and -1 if not.)*

Proof. If $T \in \Lambda^k(V)$, then T is skew-symmetric and so, writing σ as the product of transpositions, we obtain $\sigma.T = (\operatorname{sgn}\sigma)T$.

The other direction follows by considering the action of transpositions on T. □

Definition 68. For $T \in T^{(0,k)}(V)$, define $\operatorname{Alt}(T) \in T^{(0,k)}(V)$ by

$$\operatorname{Alt}(T) = \frac{1}{k!}\sum_{\sigma\in S_k} \operatorname{sgn}(\sigma)(\sigma.T),$$

where $\operatorname{sgn}(\sigma)$ is the sign of the permutation σ, so it is 1 if σ is an even permutation (i.e., $\sigma \in A_k$, the alternating group) and -1 if σ is an odd permutation.

The following proposition is straightforward:

Proposition 69. *(1) If $T \in T^{(0,k)}(V)$, then* $\operatorname{Alt}(T) \in \Lambda^k(V)$.
(2) If $\omega \in \Lambda^k(V)$, then $\operatorname{Alt}(\omega) = \omega$.
(3) If $T \in T^{(0,k)}(V)$, then $\operatorname{Alt}(\operatorname{Alt}(T)) = \operatorname{Alt}(T)$.

Proof. Observe that $\operatorname{Alt}(T)$ is skew symmetric: a transposition acting on $\operatorname{Alt}(T)$ yields a factor of -1, so $\operatorname{Alt}(T) \in \Lambda^k(V)$.

If $\omega \in \Lambda^k(V)$, we have by the lemma above

$$\operatorname{Alt}(\omega) = \frac{1}{k!}\sum_{\sigma\in S_k}(\operatorname{sgn}\sigma)(\sigma.\omega) = \frac{1}{k!}\sum_{\sigma\in S_k}(\operatorname{sgn}\sigma)^2\omega = \omega.$$

The third part follows from the first and second parts. □

Now consider what we will temporarily call the algebra of contravariant tensors on V:

$$\mathcal{T}^*(V) = \bigoplus_{k=0}^{\infty} T^{(0,k)}(V).$$

(Recall that $T^{(0,0)}(V) = \mathbb{R}$.) This set comes equipped with addition and scalar multiplication, but also enjoys a multiplication, via $\lambda \otimes \omega$. We further let $\Lambda^*(V) = \bigoplus_{k=0}^{\infty} \Lambda^k(V) \subset \mathcal{T}^*(V)$.

Proposition 70. *The kernel $\mathcal{I}$ of* Alt *is a two sided ideal in $\mathcal{T}^*(V)$.*

Proof. As $\operatorname{Alt} : T^{(0,k)}(V) \to T^{(0,k)}(V)$ (i.e., Alt preserves the type of the tensor), it suffices to show that for all $f \in \mathcal{I} \cap \Lambda^k(V)$ and $g \in T^{(0,\ell)}(V)$, $f \otimes g, g \otimes f \in \mathcal{I}$, i.e., $\operatorname{Alt}(g \otimes f) = \operatorname{Alt}(f \otimes g) = 0$. The proofs are nearly identical so we show only that $\operatorname{Alt}(f \otimes g) = 0$.

Recall that

$$\operatorname{Alt}(f\otimes g)\,(v_1, \dots, v_{k+\ell}) = \frac{1}{(k+\ell)!}\sum_{\sigma\in S_{k+\ell}} \operatorname{sgn}(\sigma) f(v_{\sigma(1)}, \dots, v_{\sigma(k)})g(v_{\sigma(k+1)}, \dots, v_{\sigma(k+\ell)}).$$

The essential idea is to split $S_{k+\ell}$ into cosets corresponding to the subgroup fixing the last ℓ indices. Indeed, let

$$G = \{\sigma \in S_{k+\ell} \mid \sigma \text{ fixes } k+1, \dots, k+\ell\}.$$

Observe that

$$\sum_{\sigma \in G} (\sigma.(f \otimes g))(v_1, \dots, v_{k+\ell}) = \sum_{\sigma \in S_k} (\operatorname{sgn} \sigma)(\sigma.f)(v_1, \dots, v_k) g(v_{k+1}, \dots, v_{k+\ell}).$$

Now, for $\sigma_0 \in S_{k+\ell} \setminus G$, we let

$$\sigma_0 G = \{\sigma_0 \sigma' \mid \sigma' \in G\}$$

denote the coset of G containing σ_0. For such a coset, we have

$$\sum_{\sigma \in \sigma_0 G} (\sigma.(f \otimes g)) = (\operatorname{sgn} \sigma_0) \left(\sigma_0. \left(\sum_{\sigma' \in G} (\operatorname{sgn} \sigma')(\sigma'.(f \otimes g)) \right) \right) = (\operatorname{sgn} \sigma_0) \left(\sigma_0. \left(\sum_{\sigma' \in S_k} (\operatorname{sgn} \sigma')(\sigma'.f) \otimes g \right) \right).$$

We let $S_{k+\ell}/G$ denote the set of cosets and now write

$$\begin{aligned}
\sum_{\sigma \in S_{k+\ell}} (\sigma.(f \otimes g)) &= \sum_{\sigma_0 \in S_{k+\ell}/G} \sum_{\sigma \in \sigma_0 G} (\sigma.(f \otimes g)) \\
&= \sum_{\sigma_0 \in S_{k+\ell}/G} (\operatorname{sgn} \sigma_0) \left(\sigma_0. \left(\sum_{\sigma' \in S_k} (\operatorname{sgn} \sigma')(\sigma'.f) \otimes g \right) \right) \\
&= k! \sum_{\sigma_0 \in S_{k+\ell}/G} (\operatorname{sgn} \sigma_0) \left(\sigma_0. \left(\operatorname{Alt}(f) \otimes g \right) \right) = 0
\end{aligned}$$

with the last equality holding because $f \in \ker \operatorname{Alt}$. □

As a consequence of the proposition, learn that $\Lambda^*(V) = \mathcal{T}^*(V)/\mathcal{I}$ is an associative algebra; in particular, it inherits a binary operation $\wedge$ arising from $\otimes$. This product is uniquely defined up to a (grading-dependent) constant factor: there must be constants $c_{k,\ell}$ so that if $\omega \in \Lambda^k(V)$ and $\eta \in \Lambda^\ell(V)$,

$$\omega \wedge \eta = c_{k,\ell} \operatorname{Alt}(\omega \otimes \eta).$$

We therefore define the wedge product by specifying these two constants. There are at least two reasonable definitions for $\omega \in \Lambda^k(V)$ and $\eta \in \Lambda^\ell(V)$:

(1) Both Lee and Spivak use the convention $\omega \wedge \eta = \dfrac{(k+\ell)!}{k!\ell!} \operatorname{Alt}(\omega \otimes \eta)$, while
(2) Bourbaki uses the convention $\omega \wedge \eta = \operatorname{Alt}(\omega \otimes \eta)$.

If $e_1, \dots, e_n$ is a basis for V and $\epsilon^1, \dots, \epsilon^n$ the corresponding dual basis for V^*, then Lee's convention yields that

$$\left(\epsilon^1 \wedge \dots \wedge \epsilon^n\right)(e_1, \dots, e_n) = 1,$$

the volume of the unit parallelopiped. Bourbaki's convention yields $1/n!$, i.e., the volume of the unit simplex. Even though the second convention might seem more natural, we'll stick with Lee's convention; it turns out that this makes various formulas later much simpler partly because we typically define integration over cubes/parallelopipeds rather than simplices.

Observe that the wedge product is *graded symmetric*, i.e., if $\omega \in \Lambda^k(V)$ and $\eta \in \Lambda^\ell(V)$, then

$$\eta \wedge \omega = (-1)^{k\ell} \omega \wedge \eta.$$

Theorem 71. *If $v_1, \dots, v_n$ is a basis for V and $\lambda^1, \dots, \lambda^n$ the corresponding dual basis for V^*, then the set*

$$\left\{\lambda^{i_1} \wedge \lambda^{i_2} \wedge \dots \wedge \lambda^{i_k} \mid 1 \leq i_1 < i_2 < \dots < i_k \leq n\right\}$$

is a basis for $\Lambda^k(V)$ and so

$$\dim \Lambda^k(V) = \binom{n}{k}.$$

Proof. Take a representative in $T^{(0,k)}(V)$ for $\omega \in \Lambda^k(V)$ and write (in a mild abuse of notation)

$$\omega = \sum_{1 \leq i_1, \dots, i_k \leq n} a_{i_1 \dots i_k} \lambda^{i_1} \otimes \dots \otimes \lambda^{i_k}.$$

As $\omega \in \Lambda^k(V)$, $\operatorname{Alt}(\omega) = \omega$, so

$$\omega = \sum_{1 \leq i_1, \dots, i_k \leq n} a_{i_1 \dots i_k} \operatorname{Alt}\left(\lambda^{i_1} \otimes \dots \otimes \lambda^{i_k}\right).$$

Now we observe that

$$\operatorname{Alt}\left(\lambda^{i_1} \otimes \dots \otimes \lambda^{i_k}\right) = \begin{cases} 0 & \text{some index is repeated} \\ \frac{\operatorname{sgn}(\sigma)}{k!} \lambda^{\sigma(i_1)} \wedge \dots \wedge \lambda^{\sigma(i_k)} & \sigma \text{ is the permutation ordering the } i_j \end{cases}$$

so that the claimed elements span $\Lambda^k(V)$.

To see that they are linearly independent, we apply them to basis elements and see that for $1 \leq i_1 < \dots < i_k \leq n$ and $1 \leq j_1 < \dots j_k \leq n$,

$$\left(\lambda^{i_1} \wedge \dots \lambda^{i_k}\right)\left(e_{j_1}, \dots, e_{j_k}\right) = \begin{cases} 1 & i_r = j_r \text{ for all } r, \\ 0 & \text{otherwise.} \end{cases}$$

□

The above theorem tells us that if $\dim(V) = n$, then $\dim \Lambda^n(V) = 1$. Moreover, if $L : V \to V$ is an endomorphism, it induces a map $L^* : \Lambda^n(V) \to \Lambda^n(V)$; as an endomorphism of a one-dimensional vector space, it must be given by multiplication by a constant. What is this constant? We turn to the example when $V = \mathbb{R}^n$.

Lemma 72. *Let $V = \mathbb{R}^n$ equipped with the standard basis and let $L : V \to V$ be a linear transformation. If A is the matrix representation of L with respect to the standard basis, then the induced map $L^* : \Lambda^n(V) \to \Lambda^n(V)$ is given by multiplication by* $\det A$.

Proof. Let $e_1, \dots, e_n$ denote the standard basis on $\mathbb{R}^n$ and let $\epsilon^1, \dots, \epsilon^n$ denote the corresponding dual basis for $(\mathbb{R}^n)^*$. A basis for $\Lambda^n(\mathbb{R}^n)$ is given by the single element $\omega = \epsilon^1 \wedge \dots \wedge \epsilon^n$. We must therefore write $L^*\omega$ in terms of ω; to do this it suffices to evaluate $L^*\omega$ on the basis for $\mathbb{R}^n$. In what follows we use a_j to denote the j-th column of the matrix

A and a_j^i to denote the i-th entry of a_j.

$$\begin{aligned}
L^*\omega(e_1,\dots,e_n) &= \omega\left(Le_1,\dots,Le_n\right) \\
&= \omega\left(a_1,\dots,a_n\right) \\
&= \omega\left(\sum_{i=1}^n a_1^i e_i,\dots,\sum_{i=1}^n a_n^i e_i\right) \\
&= \sum_{i_1,\dots,i_n=1}^n a_1^{i_1}a_2^{i_2}\dots a_n^{i_n}\omega\left(e_{i_1},\dots,e_{i_n}\right).
\end{aligned}$$

As ω is alternating, the only potentially non-vanishing terms in the sum are those for which all i_k are distinct, i.e., those arising from a permutation σ of $1,\dots,n$. We further use our characterization of the action of σ on alternating forms to write

$$\begin{aligned}
L^*\omega\left(e_1,\dots,e_n\right) &= \sum_{\sigma\in S_n} a_1^{\sigma(1)}a_2^{\sigma(2)}\dots a_n^{\sigma(n)}\omega\left(e_{\sigma(1)},\dots,e_{\sigma(n)}\right) \\
&= \sum_{\sigma\in S_n} a_1^{\sigma(1)}a_2^{\sigma(2)}\dots a_n^{\sigma(n)}(\operatorname{sgn}\sigma)\omega(e_1,\dots,e_n) \\
&= \sum_{\sigma\in S_n}(\operatorname{sgn}\sigma)a_1^{\sigma(1)}a_2^{\sigma(2)}\dots a_n^{\sigma(n)} = \det A.
\end{aligned}$$

□

Corollary 73. *If $v_1,\dots,v_n$ is a basis for V and $\omega\in\Lambda^n(V)$, and $w_j=\sum_{i=1}^n a_j^i v_i$, then*

$$\omega(w_1,\dots,w_n)=\det(a_j^i)\omega(v_1,\dots,v_n).$$

9.4. **Differential forms and the Frobenius theorem.** We now put these on bundles as we did before to get a bundle $\Lambda^k(TM)$. We define $\Omega^k(M)$, the space of differential forms of degree k, to be the space of smooth sections of the bundle $\Lambda^k(TM)$. The space of differential forms is $\Omega^*(M)=\bigoplus_k\Omega^k(M)$. Note that $\Omega^0(M)$ is canoncially identified with $C^\infty(M)$. The wedge product extends trivially to sections and we obtain an associative bilinear operation $\Omega^k(M)\times\Omega^\ell(M)\to\Omega^{k+\ell}(M)$ given by $(\omega,\eta)\mapsto\omega\wedge\eta$, where $(\omega\wedge\eta)_p=\omega_p\wedge\eta_p$.

As a smooth map $F:M\to N$ induces a map of tangent spaces, it therefore induces maps $F^*:\Lambda^k(T_{F(p)}N)\to\Lambda^k(T_pM)$ for each $p\in M$ and we get a pullback of sections $F^*\omega\in\Omega^k(M)$ for any $\omega\in\Omega^k(N)$.

Proposition 74. *The wedge product is bilinear, graded symmetric, and $F^*(\omega\wedge\eta)=(F^*\omega)\wedge(F^*\eta)$.*

Proof. The only part not immediately obvious from the properties at each point is the pullback statement. To see that they agree, we verify it at each point. Observe that $F^*(\omega\otimes\eta)=(F^*\omega)\otimes(F^*\eta)$ for any $\omega\in T^{(0,k)}(T_pM)$ and $\eta\in T^{(0,\ell)}(T_pM)$. We then recall that

$$\omega\wedge\eta=\frac{(k+\ell)!}{k!\ell!}\operatorname{Alt}(\omega\otimes\eta)$$

and so, as F^* distributes over the tensor product in each of the terms defining Alt,

$$F^*(\omega\otimes\eta)=F^*\operatorname{Alt}(\omega\otimes\eta)=\operatorname{Alt}(F^*\omega\otimes F^*\eta)=(F^*\omega)\wedge(F^*\eta)$$

□

We'd now like to take the differential of a k-form. In an open set $U \subset \mathbb{R}^n$ (i.e., in a chart on M), for $\omega \in \Omega^k(U)$, we define $d\omega \in \Omega^{k+1}(U)$ by using that we've already defined $df \in \Omega^1(M)$ for $f \in \Omega^0(M) = C^\infty(M)$. Indeed, recall that for a smooth function f we can write

$$df = \sum_{j=1}^{n} \frac{\partial f}{\partial x^j} dx^j.$$

Now, if we write

$$\omega = \sum_I \omega_I(x)\, dx^I,$$

we define

$$d\omega = \sum_I d\omega_I \wedge dx^I,$$

where $d\omega^I \in \Omega^1(U)$ is the differential of the smooth function $\omega_I(x)$.

We claim that this is independent of the coordinate system and therefore pieces together to provide a map $d : \Omega^k(M) \to \Omega^{k+1}(M)$. We could check this directly (by changing coordinates, seeing how the coefficients change under the basis change, using the chain rule, and verifying that the factors of the Jacobian all work out), but instead we'll be sneaky.

Proposition 75. *The map $d : \Omega^k(U) \to \Omega^{k+1}(U)$ defined above satisfies*

(1) $d(\omega_1 + \omega_2) = d\omega_1 + d\omega_2$,
(2) For $\omega \in \Omega^k(U)$ and $\eta \in \Omega^\ell(U)$,$d(\omega \wedge \eta) = (d\omega) \wedge \eta + (-1)^k \omega \wedge d\eta$, and
(3) $d(d\omega) = 0$.

Proof. The first statement is easy. The second is also straightforward: take $\omega = f dx^I$ and $\eta = g dx^J$ (linearity means this suffices), then

$$\begin{aligned}
\omega \wedge \eta &= (fg) dx^I \wedge dx^J, \\
d(\omega \wedge \eta) &= g\, df \wedge dx^I \wedge dx^J + f\, dg \wedge dx^I \wedge dx^J \\
&= g(d\omega \wedge dx^J) + (-1)^k f\, dx^I \wedge dg \wedge dx^J \\
&= d\omega \wedge \eta + (-1)^k \omega \wedge d\eta.
\end{aligned}$$

Finally, for the third statement, it's again enough to check it for k-forms of the form $\omega = f\, dx^I$, so that

$$\begin{aligned}
d\omega = df \wedge dx^I &= \sum_\alpha \frac{\partial f}{\partial x^\alpha} \wedge dx^I, \\
d(d\omega) &= \sum_{\alpha,\beta} \frac{\partial^2 f}{\partial x^\alpha \partial x^\beta} dx^\alpha \wedge dx^\beta \wedge dx^I.
\end{aligned}$$

In the last sum, the terms with $\alpha = \beta$ vanish because $dx^i \wedge dx^i = 0$. If $\alpha \neq \beta$, each term appears twice, but with differing signs, because $dx^\alpha \wedge dx^\beta = -dx^\beta \wedge dx^\alpha$, and so the terms vanish. ∎

We now claim that these desired properties (and the action on functions) characterize d.

Proposition 76. *Suppose $\tilde{d} : \Omega^k(U) \to \Omega^{k+1}(U)$ satisfies the following for all k and all $\omega \in \Omega^k(U)$ and $\eta \in \Omega^\ell(U)$:*

(1) $\tilde{d}(\omega + \eta) = \tilde{d}\omega + \tilde{d}\eta$,

(2) $\tilde{d}(\omega \wedge \eta) = \tilde{d}\omega \wedge \eta + (-1)^k \omega \wedge \tilde{d}\eta$,
(3) $\tilde{d}(\tilde{d}\omega) = 0$, *and*
(4) $\tilde{d}f = df$.

Then $\tilde{d} = d$.

Proof. The linearity of d and $\tilde{d}$ imply that it suffices to show that $\tilde{d}\omega = d\omega$ for $\omega = f\,dx^I$. By the second property,

$$\tilde{d}(f\,dx^I) = \tilde{d}(f \wedge dx^I) = df \wedge dx^I + f \wedge \tilde{d}(dx^I),$$

so it now suffices to show that $\tilde{d}(dx^I) = 0$. By the second property it is enough to show that $\tilde{d}(dx^i) = 0$, but $dx^i = \tilde{d}x^i$ by the fourth property and so by the third property $\tilde{d}(dx^i) = 0$. □

Corollary 77. *There is a unique* $d : \Omega^k(M) \to \Omega^{k+1}(M)$ *satisfying the four properties above.*

Here is an invariant definition (we probably won't prove this in class):

Theorem 78. *If* $\omega \in \Omega^k(M)$, *there is a unique* $(k+1)$*-form* $d\omega \in \Omega^{k+1}(M)$ *so that for all smooth vector fields* $X_1, \dots, X_{k+1} \in \mathcal{X}(M)$,

$$\begin{aligned} d\omega\left(X_1, \dots, X_{k+1}\right) = &\sum_{j=1}^{k+1} X_j \left(\omega(X_1, \dots, \hat{X}_j, \dots, X_{k+1})\right) \\ &+ \sum_{1 \leq i < j \leq k+1} (-1)^{i+j} \omega\left([X_i, X_j], X_1, \dots, \hat{X}_i, \dots, \hat{X}_j, \dots, X_{k+1}\right). \end{aligned}$$

where the notation $\hat{X}_j$ *here means that* X_j *has been omitted from the list.*

We'll instead prove the following simpler description of d:

Proposition 79. *If* $\omega \in \Omega^1(M)$ *and* $X, Y \in \mathcal{X}(M)$, *then*

$$d\omega(X,Y) = X(\omega(Y)) - Y(\omega(X)) - \omega([X,Y]).$$

Proof. Here is a deeply unsatisfying[9] proof: write $\omega = \sum_j \omega_j dx^j$, $X = \sum_i X^i \frac{\partial}{\partial x^i}$, $Y = \sum_i Y^i \frac{\partial}{\partial x^i}$, then compute:

$$\begin{aligned} d\omega(X,Y) &= \sum_{i,j} \frac{\partial \omega_j}{\partial x^i}\left(X^i Y^j - X^j Y^i\right), \\ X(\omega(Y)) - Y(\omega(X)) &= \sum_{i,j} \left(X^i \frac{\partial}{\partial x^i}(\omega_j Y^j) - Y^i \frac{\partial}{\partial x^i}(\omega_j X^j)\right), \\ \omega([X,Y]) &= \sum_{i,j} \omega_j \left(X^i \frac{\partial Y^j}{\partial x^i} - Y^i \frac{\partial X^j}{\partial x^i}\right). \end{aligned}$$

Using the Leibniz rule finishes the proof. □

This lets us reformulate the Frobenius theorem:

Theorem 80 (Frobenius, take 2)**.** *Suppose* $\Delta \subset TM$ *is a* k*-dimensional distribution. Let* $\mathcal{I} \subset \Omega^*(M)$ *denote the annihilator of* Δ, *i.e.,* $\omega \in \mathcal{I}$ *if and only if* $\omega|_\Delta = 0$. *The distribution* Δ *is integrable if and only if* $d(\mathcal{I}) \subset \mathcal{I}$.

[9] A more satisfying proof takes us a bit too far afield right now.

Proof. It's enough to show this on the 1-forms that generate $\mathcal{I}$. Suppose $\omega \in \mathcal{I}$ is a 1-form, then

$$d\omega(X, Y) = X(\omega(Y)) - Y(\omega(X)) - \omega([X, Y])$$

by the proposition above.

For the forward direction, if Δ is integrable then all terms on the right vanish when X, Y are sections of Δ, so $d\omega \in \mathcal{I}$. In the other direction, one sees then that $\omega([X, Y]) = 0$ for all $\omega \in \mathcal{I}$. As $[X, Y]$ is annihilated by the entire annihilator, we must in fact have $[X, Y] \in \Delta$, so Δ is integrable. □

To conclude this section, we include how the differential interacts with the pullback.

Proposition 81. *If $F : M \to N$ is smooth and $\omega \in \Omega^k(N)$, then*

$$F^*(d\omega) = d(F^*\omega).$$

Proof. By linearity, we can assume $\omega = g dx^I$ in a coordinate chart. We proceed by induction. For $k = 0$, we have, for $v \in T_pM$

$$\begin{aligned} F^*(dg)_p(v) &= dg(DF_p v) = (DF)_p(v)(g) = v(g \circ F) \\ &= d(g \circ F)(v) = d(F^*g)(v), \end{aligned}$$

so $F^*(dg) = d(F^*g)$. (We've done this computation before.)

Now assume the claim is true for $(k-1)$-forms. If $\omega = g\, dx^{i_1} \wedge \cdots \wedge dx^{i_k}$, we have

$$\begin{aligned} d(F^*\omega) &= d\left(F^*\left(g\, dx^{i_1} \wedge \cdots \wedge dx^{i_{k-1}}\right) \wedge F^* dx^{i_k}\right) \\ &= d\left(F^*\left(g\, dx^{i_1} \wedge \cdots \wedge dx^{i_{k-1}}\right) + d(x^{i_k} \circ F)\right) \\ &= F^*\left(d\left(g\, dx^{i_1} \wedge \cdots \wedge dx^{i_{k-1}}\right)\right) \wedge F^* dx^{i_k} + 0 \\ &= F^*\left(dg \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_{k-1}}\right) \wedge F^* dx^{i_k} \\ &= F^*\left(dg \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_k}\right) = F^*(d\omega), \end{aligned}$$

where the 0 appears because $d(d(x^{i_k} \circ F)) = 0$. □

10. Integration and cohomology

Our aim in this section is to describe how to integrate differential forms over submanifolds. The two stars of this section are Stokes's theorem and de Rham cohomology. For a differential form $\omega \in \Omega^k(M)$ and a submanifold $N \subset M$ of dimension $k+1$, Stokes's theorem relates the following two integrals (which we'll define):

$$\int_N d\omega = \int_{\partial N} \omega.$$

De Rham cohomology is concerned with the groups

$$H^k_{\mathrm{dR}}(M) = \ker(d_k)/\operatorname{Im}(d_{k-1}),$$

where $d_k : \Omega^k(M) \to \Omega^{k+1}(M)$. We'll see that these two theorems are related.

Forms in the kernel of d_k are called *closed* k-forms, while those in the image of d_{k-1} are called *exact* k-forms.

10.1. **Integrating differential forms.** We'd now like to integrate differential k-forms over k-dimensional submanifolds. If $\omega = f dx^1 \wedge \cdots \wedge dx^k$ in $\mathbb{R}^k$, we want to use that we already know how to integrate and define

$$\int_{[0,1]^k} \omega = \int_{[0,1]^k} f dx^1 \wedge \cdots \wedge dx^k = \int_{[0,1]^k} f\, dx^1 \ldots dx^k.$$

The key point of dealing with forms is that we'd like the integral to be independent of how we parametrize the domain.

A motivating example is of line integrals; integrating a function behaves poorly when reparametrizing the curve (as you learned in a multivariable calculus class!). Instead, we use a 1-form ω, which has a sort of directionality built in.

Definition 82. Suppose $\gamma : [0,1] \to M$ is a curve and $\omega \in \Omega^1(M)$. We define the integral of ω over γ by

$$\int_\gamma \omega = \int_0^1 \gamma^* \omega.$$

We claim that this integral is independent of the (orientation-preserving) parametrization of the interval $[0,1]$. Fix a coordinate system and write $\omega = \sum \omega_i dx^i$ and $\gamma(t) = (\gamma^1(t), \ldots, \gamma^n(t))$. We then have

$$(\gamma^*\omega)_t = \sum_{\omega_i} (\gamma(t)) d(\gamma^i(t)) = \sum (\omega_i \circ \gamma)(t) \frac{d\gamma^i}{dt}\, dt.$$

If $p : [0,1] \to [0,1]$ is a smooth, increasing, diffeomorphism, then

$$\begin{aligned} \int_{\gamma \circ p} \omega = \int_0^1 (\gamma \circ p)^* \omega &= \sum \int_0^1 (\omega_i \circ \gamma \circ p)(t) \frac{d\gamma^i}{dp} \frac{dp}{dt}\, dt \\ &= \sum \int_0^1 (\omega_i \circ \gamma)(u) \frac{d\gamma^i}{du}\, du = \int_\gamma \omega \end{aligned}$$

by the chain rule.

As an aside, the above computation indicates that the objects you could integrate over curves should really be 1-*densities*, i.e., things that transform like the absolute value of the derivative of the reparametrization. Even though this approach would allow us to avoid issues of orientation, we avoid it here because the differential of such an object is not nearly as nice.

More generally, we say that the standard k-cube is $I^k = [0,1]^k \subset \mathbb{R}^k$.

Definition 83. A (smooth) singular k-cube in M is a (smooth) map $c : I^k \to M$.

Definition 84. For $\omega \in \Omega^k(M)$ and a singular k-cube c, we define

$$\int_c \omega = \int_{I^k} c^* \omega.$$

Since $c^*\omega \in \Omega^k(I^k)$, which is one-dimensional, we can write $c^*\omega = f\, dx^1 \wedge \cdots \wedge dx^k$ and so we already know how to integrate $c^*\omega$ over I^k.

An exercise using the change of variables formula for integrals shows that $\int_c \omega$ is independent of orientation-preserving reparametrization of c:

Proposition 85. *Let $c : [0,1]^n \to \mathbb{R}^n$ be an invertible singular cube with $\det \frac{\partial c^i}{\partial x^j} \geq 0$ on $[0,1]^n$. If $\omega \in \Omega^n(\mathbb{R}^n)$ be given by $\omega = f\,dx^1 \wedge \cdots \wedge dx^n$, then*

$$\int_c \omega = \int_{c([0,1]^n)} f.$$

Proof. We compute:

$$\begin{aligned}\int_c \omega = \int_{[0,1]^n} c^*\omega &= \int_{[0,1]^n} f(c(x))(\det \frac{\partial c}{\partial x}) dx^1 \wedge \cdots \wedge dx^n \\ &= \int_{[0,1]^n} f(c(x)) \left|\det \frac{\partial c}{\partial x}\right| dx^1 \wedge \cdots \wedge dx^n = \int_{c([0,1]^n)} f.\end{aligned}$$

□

Corollary 86. *Let $p : [0,1]^k \to [0,1]^k$ be one-to-one, onto, smooth, and $\det p' \geq 0$. Let c be a singular k-cube in M, and $\omega \in \Omega^k(M)$. Then*

$$\int_c \omega = \int_{c \circ p} \omega.$$

Proof. Again we compute:

$$\int_{c\circ p} \omega = \int_{I^k} p^* c^* \omega = \int_{I^k} c^*\omega = \int_c \omega.$$

□

Here is another variation on the theme:

Proposition 87. *Suppose M is an oriented n-manifold and $c_1, c_2 : [0,1]^n \to M$ are singular n-cubes that can be extended to diffeomorphisms in a neighborhood of $[0,1]^n$. Assume both are orientation-preserving. If $\omega \in \Omega^n(M)$ is supported in $c_1([0,1]^n) \cap c_2([0,1]^n)$, then*

$$\int_{c_1} \omega = \int_{c_2} \omega.$$

Proof. The main idea is that you can write

$$c_1 = c_1 \circ (c_1^{-1} \circ c_2)$$

on the support of ω. □

The above proposition lets us integrate n-forms on oriented manifolds: we cover M by open sets U_i each contained in $c_i([0,1]^n)$ and use a partition of unity. This works as long as the support of ω is compact (so that the sum is finite). Note that the answer depends on orientation! Reversing the orientation flips the sign of the integral.

10.2. **Another excursion into the world of boundaries.** As our aim is Stokes's theorem, we need to grapple somewhat with the notion of boundaries. We start by defining a singular chain:

Definition 88. A singular k-chain is a formal sum of singular k-cubes:

$$\sum_{j=1}^{k} a_i c_i,$$

where $a_i \in \mathbb{R}$ and c_i is a singular k-cube. For a singular k-chain $\sum_i a_i c_i$ and a k-form ω, we define

$$\int_{\sum a_i c_i} \omega = \sum_i a_i \int_{c_i} \omega = \sum_i a_i \int_{I^k} c_i^* \omega.$$

We can think of elements of Ω^k as linear functionals on the space of k-chains.

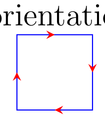
We now define the "boundary operator" ∂ from the space of k-chains to the space of $(k-1)$-chains by defining it on cubes and extending linearly. For a singular k-cube c, we define

$$\partial c = \sum_{i=1}^{k} \sum_{\alpha=0,1} (-1)^{i+\alpha} c_{(i,\alpha)},$$

where $c_{(i,\alpha)} : I^{k-1} \to M$ is defined by putting α into the i-th slot of c, i.e.,

$$c_{(i,\alpha)}(t^1, \dots, t^{k-1}) = c(t^1, \dots, t^{i-1}, \alpha, t^i, \dots, t^{k-1}).$$

Three quick examples:

(1) If $c : I \to M$ is a singular 1-cube, then $\partial c = c(1) - c(0)$ (remember this is a formal sum!).
(2) If $c : I^2 \to M$ is a singular 2-cube, then $\partial c = c(1, \cdot) - c(0, \cdot) - c(\cdot, 1) + c(\cdot, 0)$. You can think of the signs as giving you orientations on each piece. In this example it gives you a schematic picture like this: 
(3) If $c : \{0\} \to M$ is a 0-cube, then $\partial c = 1$ (this is sort of a convention).

Proposition 89. *If c is any singular k-chain on M, then $\partial(\partial c) = 0$ (i.e., $\partial^2 = 0$).*

Proof. It suffices to check this for a singular k-cube c. We'll start by considering the map $\mathrm{Id}_k : [0,1]^k \to [0,1]^k$ with $\mathrm{Id}_k(x^1, \dots, x^k) = (x^1, \dots, x^k)$.

Suppose $1 \leq i \leq j \leq k-1$ and $\alpha, \beta \in \{0,1\}$. Because $i \leq j$, by carefully writing down the two maps, you can see that

$$\left(\mathrm{Id}^k_{(i,\alpha)}\right)_{(j,\beta)} = \left(\mathrm{Id}^k_{(j+1,\beta)}\right)_{(i,\alpha)}.$$

Applying this observation now to a general singular k-cube shows that for $1 \leq i \leq j \leq k-1$,

$$(c_{(i,\alpha)})_{(j,\beta)} = (c_{(j+1,\beta)})_{(i,\alpha)},$$

so that

$$\begin{aligned}
\partial(\partial c) &= \partial \left(\sum_{i=1}^{k} \sum_{\alpha=0,1} (-1)^{i+\alpha} c_{(i,\alpha)} \right) \\
&= \sum_{i=1}^{k} \sum_{\alpha=0,1} \sum_{j=1}^{k-1} \sum_{\beta=0,1} (-1)^{i+j+\alpha+\beta} (c_{(i,\alpha)})_{(j,\beta)}.
\end{aligned}$$

We now split the sum into two terms: those where $i \leq j$ and those where $i > j$:

$$\begin{aligned}\partial(\partial c) &= \sum_{i=1}^{k}\sum_{j=i}^{k-1}\sum_{\alpha,\beta=0,1}(-1)^{i+j+\alpha+\beta}(c_{(i,\alpha)})_{(j,\beta)} + \sum_{i=1}^{k}\sum_{j=1}^{i-1}\sum_{\alpha,\beta=0,1}(-1)^{i+j+\alpha+\beta}(c_{(i,\alpha)})_{(j,\beta)} \\ &= \sum_{i=1}^{k}\sum_{j=i}^{k-1}\sum_{\alpha,\beta=0,1}(-1)^{i+j+\alpha+\beta}(c_{(i,\alpha)})_{(j,\beta)} + \sum_{i=1}^{k}\sum_{j=1}^{i-1}\sum_{\alpha,\beta=0,1}(-1)^{i+j+\alpha+\beta}(c_{(j,\beta)})_{(i-1,\alpha)},\end{aligned}$$

where the second equality follows from our earlier observation. We now note that the first sum is empty when $i = k$ and the second sum is empty when $i = 1$, so we can reindex the second sum using $i - 1$ and write

$$\begin{aligned}\partial(\partial c) &= \sum_{i=1}^{k-1}\sum_{j=i}^{k-1}\sum_{\alpha,\beta=0,1}(-1)^{i+j+\alpha+\beta}(c_{(i,\alpha)})_{(j,\beta)} + \sum_{i=1}^{k-1}\sum_{j=1}^{i}\sum_{\alpha,\beta=0,1}(-1)^{i+1+j+\alpha+\beta}(c_{(j,\beta)})_{(i,\alpha)} \\ &= \sum_{i=1}^{k-1}\sum_{j=i}^{k-1}\sum_{\alpha,\beta=0,1}(-1)^{i+j+\alpha+\beta}(c_{(i,\alpha)})_{(j,\beta)} + \sum_{i=1}^{k-1}\sum_{j=i}^{k-1}\sum_{\alpha,\beta=0,1}(-1)^{i+j+\alpha+\beta+1}(c_{(i,\alpha)})_{(j,\beta)} = 0.\end{aligned}$$

where the second equality follows from exchanging the order of the summation in the second term (and reindexing). □

10.2.1. *Induced orientations on the boundary.* Suppose M is an oriented manifold with boundary ∂M. Let $p \in \partial M$. We way that a vector $v \in T_pM$ is "outward pointing" if in any chart (x, U) of p to $\mathbb{H}^n$, $(Dx)_p(v)$ is "outward pointing" (i.e., it points to the outside of $\mathbb{H}^n$).

If μ is an orientation on M (i.e., an orientation of the bundle TM), then it induces an orientation $\partial\mu$ on ∂M by

$$[v_1, \dots, v_{n-1}] \in (\partial\mu)_p \quad \text{if and only if} \quad [w, v_1, \dots, v_{n-1}] \in \mu_p$$

for all outward pointing $w \in T_pM$.

As a quick warning, this is $(-1)^n$ times the standard orientation on $\partial\mathbb{H}^n = \mathbb{R}^{n-1}$. The reason for this discrepancy comes from our convention for ∂ on chains; $c(n, 0)$ has the sign $(-1)^n$.

10.3. **Stokes's theorem.** First we prove a simpler version of Stokes's theorem (for k-cubes):

Theorem 90. *If ω is a $k-1$-form on M and c is a singular k-cube, then*

$$\int_c d\omega = \int_{\partial c} \omega.$$

Proof. This follows from the fundamental theorem of calculus. By the definition of our integral in terms of the pullback, it suffices to prove it on I^k. We therefore suppose $\omega \in \Omega^{k-1}(I^k)$. By linearity, it suffices to consider ω of the form

$$\omega = f\, dx^1 \wedge \cdots \wedge \widehat{dx^j} \wedge \cdots \wedge dx^k,$$

where the "hat" notation means that dx^j is omitted from the wedge product. We then compute:

$$\int_{\partial I^k} \omega = \sum_{i=1}^{k}\sum_{\alpha=0,1}(-1)^{j+\alpha}\int_{I^{k-1}}(\mathrm{Id}^k_{(i,\alpha)})^*\omega.$$

By the definition of these maps, we have (using t coordinates on I^{k-1})

$$(\mathrm{Id}^k_{(i,\alpha)})^*\omega = f(t^1,\dots,t^{i-1},\alpha,t^i,\dots,t^{k-1})d(x^1\circ \mathrm{Id}^k_{(i,\alpha)})\wedge\dots\wedge d(\widehat{x^j\circ \mathrm{Id}^k_{(i,\alpha)}})\wedge\dots\wedge d(x^k\circ\mathrm{Id}^k_{(i,\alpha)})$$
$$= \begin{cases} 0 & i\neq j,\\ f(t^1,\dots,t^{j-1},\alpha,t^j,\dots,t^k)dt^1\wedge\dots\wedge dt^{k-1} & i=j,\end{cases}$$

where the second identity follows from the observation that

$$x^r\circ \mathrm{Id}^k_{(i,\alpha)} = \begin{cases} t^r & r<i,\\ \alpha & r=i,\\ t^{r-1} & r>i,\end{cases}$$

so that $d(x^i\circ\mathrm{Id}^k_{(i,\alpha)})=0$.

We therefore conclude that

$$\begin{aligned}\int_{\partial I^k}\omega &= \sum_{i=1}^k\sum_{\alpha=0,1}(-1)^{i+\alpha}\int_{I^{k-1}}(\mathrm{Id}^k_{(i,\alpha)})^*\left(f\,dx^1\wedge\dots\wedge\widehat{dx^j}\wedge\dots\wedge dx^k\right)\\ &= (-1)^{1+j}\int_{I^{k-1}} f(t^1,\dots,t^{j-1},1,t^i,\dots,t^{k-1})dt^1\wedge\dots\wedge dt^{k-1}\\ &\quad + (-1)^j\int_{I^{k-1}} f(t^1,\dots,t^{j-1},0,t^j,\dots,t^{k-1})dt^1\wedge\dots\wedge dt^{k-1}.\end{aligned}$$

We finally observe that

$$d\omega = \frac{\partial f}{\partial x^j}dx^j\wedge dx^1\wedge\dots\wedge\widehat{dx^j}\wedge\dots\wedge dx^k = (-1)^{j-1}\frac{\partial f}{\partial x^j}dx^1\wedge\dots\wedge dx^k,$$

so that

$$\int_{I^k}d\omega = \int_{[0,1]^k}(-1)^{j-1}\frac{\partial f}{\partial x^j}dx^1\wedge\dots\wedge dx^k.$$

Now writing this integral as an iterated integral (with the j-th one first) and applying the fundamental theorem of calculus yields the result. $\square$

Theorem 91 (Stokes's theorem)**.** *If M is an oriented n-manifold with boundary ∂M, ∂M is equipped with the induced orientation, and $\omega\in\Omega^{n-1}(M)$ has compact support, then*

$$\int_{\partial M}\omega = \int_M d\omega.$$

Proof. Suppose first that there is an orientation-preserving $c:[0,1]^n\to M\setminus\partial M$ so that $\operatorname{supp}\omega$ is a subset of the interior of $c([0,1]^n)$. We can then use the previous theorem to see that

$$\int_M d\omega = \int_c d\omega = \int_{\partial c}\omega = 0$$

because $\operatorname{supp}\omega$ is in the interior of $c([0,1]^n)$. Now $\omega\equiv 0$ on ∂M, so $\int_{\partial M}\omega = 0$.

Now suppose $c:[0,1]^n\to M$ is a singular n-cube so that

$$\partial M\cap c([0,1]^n) = c_{(n,0)}([0,1]^{n-1}),$$

and that $\operatorname{supp}\omega \subset c([0,1]^n)$. We then have, by our previous theorem,

$$\int_M d\omega = \int_c d\omega = \int_{\partial c} \omega = \int_{(-1)^n c_{(n,0)}} \omega = \int_{\partial M} \omega.$$

Now, in general, we take an open cover $\mathcal{O}$ of M so that for each each $O \in \mathcal{O}$, O is contained in the image of a singular k-cube c that is either

(1) disjoint from ∂M, or
(2) $\partial M \cap c([0,1]^n) = c_{(n,0)}([0,1]^{n-1})$.

We now take a partition of unity Φ subordinate to $\mathcal{O}$ and write

$$\int_M d\omega = \sum_{\phi\in\Phi} \int_M \phi\, d\omega$$

and there are only finitely many nonzero terms in the sum because $\operatorname{supp}\omega$ is compact. Now, because

$$0 = d(1) = d(\sum_{\phi\in\Phi} \phi) = \sum_{\phi\in\Phi} d\phi,$$

we observe that

$$\sum_{\phi\in\Phi} d\phi \wedge \omega = 0,$$

and so

$$\sum_{\phi\in\Phi} \phi\, d\omega = \sum_{\phi\in\Phi} d(\phi\omega).$$

As each ϕ is supported in a singular k-cube of the above form, we have

$$\begin{aligned} \int_M d\omega &= \sum_{\phi\in\Phi} \int_M d(\phi\omega) \\ &= \sum_{\phi\in\Phi} \int_{\partial M} \phi\omega = \int_{\partial M} \omega, \end{aligned}$$

proving the claim. □

As an application, we have the following:

Corollary 92. *If M is an oriented, compact manifold without boundary, then for any $\omega \in \Omega^{n-1}(M)$, we have*

$$\int_M d\omega = 0.$$

Proof. By Stokes's theorem, we have

$$\int_M d\omega = \int_{\partial M} \omega = \int_{\emptyset} \omega = 0.$$

□

10.4. **de Rham cohomology.** Suppose that M is an n-dimensional manifold. For each k, let $Z^k(M) \subset \Omega^k(M)$ denote the space of closed k-forms, i.e., for $\omega \in \Omega^k(M)$, $\omega \in Z^k$ if and only if $d\omega = 0$. Let $B^k(M) \subset \Omega^k(M)$ denote the space of exact forms, i.e., $\omega \in B^k$ if there is some $\eta \in \Omega^{k-1}(M)$ so that $\omega = d\eta$.

Observe that $B^k(M) \subset Z^k(M)$ because $d^2 = 0$.

Definition 93. The k-th de Rham cohomology group of M is

$$H^k_{\mathrm{dR}}(M) = Z^k(M)/B^k(M).$$

A similar definition applies with compact supports, too:

$$H^k_{\mathrm{c}}(M) = Z^k_{\mathrm{c}}(M)/B^k_{\mathrm{c}}(M),$$

where $Z^k_{\mathrm{c}}(M)$ is the space of closed k-forms with compact support and $B^k_{\mathrm{c}}(M)$ is the space of forms $d\eta$, where $\eta \in \Omega^{k-1}(M)$ has compact support.

Proposition 94 (Poincaré Lemma)**.** $H^k_{\mathrm{dR}}(\mathbb{R}^n) = \begin{cases} \mathbb{R} & k = 0, \\ 0 & k \neq 0. \end{cases}$

Before we prove the proposition, we need a bit of additional discussion.

Definition 95. M is smoothly contractible to a point if there is a smooth function $H : M \times [0,1] \to M$ so that $H(p,1) = p$ for all $p \in M$ and $H(p,0) = p_0$ for a fixed $p_0 \in M$. The map H is called a (smooth) homotopy.

Now, for each $t \in [0,1]$, define $\iota_t : M \to M \times [0,1]$ by $p \mapsto (p,t)$. We can write each $\omega \in \Omega^k(M \times [0,1])$ as

$$\omega = \omega_1 + dt \wedge \eta,$$

where $\eta \in \Omega^{k-1}(M)$ and $\omega_1(v_1, \dots, v_k) = 0$ if some $v_i \in \ker(\pi_M)_*$. (Here $\pi_M : M \times [0,1] \to M$ is projection onto the first factor.)

We now define an operator $I : \Omega^k(M \times [0,1]) \to \Omega^{k-1}(M)$ by

$$(I\omega)_p(v_1, \dots, v_{k-1}) = \int_0^1 \eta_{(p,t)}((\iota_t)_* v_1, \dots, (\iota_t)_* v_{k-1})\, dt.$$

Theorem 96. *For any* $\omega \in \Omega^k(M \times [0,1])$,

$$\iota_1^*\omega - \iota_0^*\omega = d(I\omega) + I(d\omega).$$

Before we prove the theorem, note that if $d\omega = 0$, then the right hand side is $d(I\omega)$, so we have the following immediate corollary.

Corollary 97. *If* $\omega \in \Omega^k(M \times [0,1])$ *is closed, then* $\iota_1^*\omega - \iota_0^*\omega \in \Omega^k(M)$ *is exact.*

Proof of theorem. Fix a coordinate chart x on M and use coordinates $(x^1, \dots, x^n, t)$ on $M \times [0,1]$. As I is linear it suffices to consider the two cases $\omega = f\, dx^I$ and $\omega = f\, dt \wedge dx^I$.

If $\omega = f\, dx^I$, we first note that $I\omega = 0$ and then write

$$d\omega = \left(\sum_{i=1}^n \frac{\partial f}{\partial x^i} dx^i \wedge dx^I \right) + \frac{\partial f}{\partial t} dt \wedge dx^I.$$

We then compute that

$$I(d\omega)_p = \left(\int_0^1 \frac{\partial f}{\partial t}(p,t)\, dt \right) dx^I = (f(p,1) - f(p,0))\, dx^I = \iota_1^*\omega - \iota_0^*\omega,$$

verifying the claim.

Now suppose $\omega = f\,dt \wedge dx^I$, so that $\iota_1^*\omega = \iota_0^*\omega = 0$. We then have that

$$I(d\omega)_p = I\left(-\sum_{i=1}^n \frac{\partial f}{\partial x^i} dt \wedge dx^i \wedge dx^I\right)_p,$$
$$d(I\omega)_p = d\left(\int_0^1 f(p,t)\,dt\right) dx^I = -I(d\omega),$$

again verifying the claim. □

Corollary 98. *If M is contractible then every closed k-form is exact for $k \geq 1$.*

Proof. Suppose $H : M \times [0,1] \to M$ is a smooth homotopy verifying that M is contractible, so $H(p,1) = p$ and $H(p,0) = p_0$. Thus $H \circ \iota_1 = \mathrm{Id}_M$ and $H \circ \iota_0 = p_0$ (the constant map).

Suppose $\omega \in \Omega^k(M)$ satisfies $d\omega = 0$. We then have

$$\omega = (H \circ \iota_1)^*\omega = \iota_1^* H^*\omega,$$

while, because $k \geq 1$,

$$0 = (H \circ \iota_0)^*\omega = \iota_0^* H^*\omega.$$

We may thus write

$$\omega = (\iota_1 - \iota_0)^*(H^*\omega) = d\left(IH^*\omega\right) - I\left(dH^*\omega\right) = d\left(IH^*\omega\right) - I(H^*d\omega) = d\left(IH^*\omega\right).$$

□

We are finally able to prove the Poincaré lemma:

Proof of Poincaré Lemma. Observe that $\mathbb{R}^n$ is contractible by the straight-line homotopy

$$H(x,t) = tx,$$

so that $H^k(\mathbb{R}^n) = 0$ for $k > 0$. Now for $k = 0$, if $f \in \Omega^0(\mathbb{R}^n)$ has $df = 0$, then f is constant by the fundamental theorem of calculus. This space of constants is $H^0_{\mathrm{dR}}(\mathbb{R}^n)$. □

Now, on $\mathbb{S}^{n-1} \subset \mathbb{R}^n$, let's define an $(n-1)$-form σ' by

$$(\sigma')_p(v_1, \ldots, v_{n-1}) = \det(p, v_1, \ldots, v_{n-1}).$$

In fact, σ' is the restriction to $\mathbb{S}^{n-1}$ of the form

$$\sigma = \sum_{i=1}^n (-1)^{i-1} x^i\, dx^1 \wedge \cdots \wedge \widehat{dx^i} \wedge \cdots \wedge dx^n.$$

Observe that $\sigma'(p) > 0$ on any positively ordered basis, so $\int_{\mathbb{S}^{n-1}} \sigma' > 0$. Let $r : \mathbb{R}^n \setminus 0 \to \mathbb{S}^{n-1}$ be the retraction $r(p) = \frac{p}{|p|}$, so that $r \circ i = \mathrm{Id}_{\mathbb{S}^{n-1}}$. (Here $i : \mathbb{S}^{n-1} \to \mathbb{R}^n \setminus 0$ denotes the inclusion map.)

Observe that $r^*\sigma'$ is closed but not exact:

$$d(r^*\sigma') = r^*(d\sigma') = 0$$

because $d\sigma' = 0$ as $d\omega'$ is an n-form but $\dim \mathbb{S}^{n-1} = n-1$. If $r^*\sigma' = d\eta$ for some η, then $\sigma' = i^*r^*\sigma' = i^*d\eta = d(i^*\eta)$, but $\int_{\mathbb{S}^{n-1}} \sigma' > 0$, so σ' cannot be exact.

Lemma 99. $(r^*\sigma')_p = \dfrac{1}{|p|^n}\sigma_p$

NOTE: I'm pausing this part of the notes for now since it looks like we might not get a chance to talk about cohomology in class. :-(

11. Lie groups

Definition 100. A topological group is a topological space G that is also group for which the maps

$$G \times G \to G, \quad (a, b) \mapsto ab,$$
$$G \to G \quad a \mapsto a^{-1}$$

are continuous.

Definition 101. A Lie group is a topological group that is also a smooth manifold on which the above maps are also smooth.

Note it's enough to check $(x, y) \mapsto xy^{-1}$ is smooth. Except in groups that are already familiar, we'll use e to denote the identity element.

Examples: $(\mathbb{R}^n, +)$, $\mathbb{S}^1 = \mathbb{R}/\mathbb{Z}$, products, $GL(n, \mathbb{R})$, $O(n)$, $SO(n)$, isometries of $\mathbb{R}^n$ (homework exercise).

Proposition 102. *If G is a topological group, then the connected component K containing the identity $e \in G$ is a closed normal subgroup of G. If G is a Lie group, then K is also a Lie group and is open.*

Proof. Note that $b \mapsto a^{-1}b$ is a homeomorphism $G \to G$, so if K is a connected component, so is $a^{-1}K$. Now, if $a \in K$, we have $e = a^{-1}a \in a^{-1}K$, we must have $a^{-1}K \subset K$ because K is the connected component containing e. As they are both connected components, $a^{-1}K = K$. This is true for all $a \in K$, so $K^{-1}K = K$, i.e., K is a subgroup.

For $b \in G$, bKb^{-1} is connected, and $e \in bKb^{-1} \cap K$, so because K is a connected component and bKb^{-1} is connected, $bKb^{-1} \subset K$, i.e., K is normal. K is closed because connected components are always closed.

If G is also a Lie group, G is locally connected and so K must also be open; K is therefore a submanifold and thus a Lie group. $\square$

11.1. **Lie algebras.** The left and right translations by $a \in G$ are the maps $G \to G$ given by

$$L_a(b) = ab,$$
$$R_a(b) = ba.$$

Both are diffeomorphisms with inverses $L_{a^{-1}}$ and $R_{a^{-1}}$, so their derivatives

$$(L_a)_* : T_bG \to T_{ab}G, \quad (R_a)_* : T_bG \to T_{ba}G$$

are isomorphisms.

Definition 103. A vector field X on G is left invariant if $(L_a)_*X = X$ for all $a \in G$, i.e.,

$$(L_a)_*X_b = X_{ab}$$

for all $a, b \in G$.

Given any $X_e \in T_eG$, there is a unique left invariant vector field X on G with value X_e at e. Indeed, let $X_a = (L_a)_*X_e$.

Proposition 104. *Left invariant vector fields are smooth.*

Proof. Let $\widetilde{X}$ be a left invariant vector field on G. It is enough to show that $\widetilde{X}$ is smooth in a neighborhood of e because L_a is a diffeomorphism and $\widetilde{X}$ is left invariant. To do this we take a smooth function f on G and show that $\widetilde{X}(f)$ is smooth.

As $\widetilde{X}_a = (L_a)_* X_e$, we note that

$$\widetilde{X}_a(f) = X_e(f \circ L_a).$$

We therefore consider the function $a \mapsto X_e(f \circ L_a)$; we claim this is a smooth function. Let $\varphi : G \times G \to G$ denote the multiplication map; this is known to be smooth because G is a Lie group.

Let (x, U) be a coordinate system around e and V be a sufficiently small neighborhood of e so that if $a, b \in V$ then $ab^{-1} \in U$. Write

$$X = \sum c^j \frac{\partial}{\partial x^j},$$

so that $c^i = X(x^i)$. At a, we have

$$\begin{aligned} c^i(a) = X_a(x^i) &= (L_a)_*(X_e)(x^i) \\ &= X_e(x^i \circ L_a). \end{aligned}$$

The claim then follows by showing that, given a smooth function f around e, the function $a \mapsto X_e(f \circ L_a)$ is smooth near e. Let $\varphi : G \times G \to G$ denote group multiplication and let i_e and i_a denote two "slice maps":

$$i_e(b) = (b, e), \quad i_a(b) = (a, b).$$

Let Y be any smooth vector field on G with $Y_e = X_e$, so that $(0, Y)$ is smooth on $G \times G$ and thus $[(0, Y)(f \circ \varphi)] \circ i_e$ is smooth on G. We now compute

$$\begin{aligned} [(0, Y)(f \circ \varphi)] \circ i_e(a) &= (0, Y)_{(a,e)}(f \circ \varphi) \\ &= 0_a(f \circ \varphi \circ i_e) + Y_e(f \circ \varphi \circ i_a) \\ &= X_e(f \circ L_a). \end{aligned}$$

□

Corollary 105. *A Lie group G always has a trivial tangent bundle (and is therefore orientable).*

Proof. Choose a basis $X_{1e}, \ldots, X_{ne}$ for T_eG and let $X_1, \ldots, X_n$ denote the corresponding left invariant vector fields. As they are linearly independent everywhere, we can define a map

$$f : TG \to G \times \mathbb{R}^n$$

by writing tangent vectors at p in terms of the X_j at p:

$$f\left(\left(p, \sum_{j=1}^{n} c^j X_{j_p}\right)\right) = \left(p, \begin{pmatrix} c^1 \\ \vdots \\ c^n \end{pmatrix}\right) \in G \times \mathbb{R}^n.$$

□

As left invariant is equivalent to L_a-related to itself for all a, we know that if X and Y are left-invariant so is $[X, Y]$.

FOR THE REST OF OUR DISCUSSION OF LIE GROUPS, X, Y denote elements of T_eG, and $\widetilde{X}$ and $\widetilde{Y}$ denote left invariant vector fields with $\widetilde{X}_e = X$ and $\widetilde{Y}_e = Y$.

Define a Lie bracket on T_eG by

$$[X, Y] = [\widetilde{X}, \widetilde{Y}]_e.$$

Definition 106. The vector space T_eG, together with the bracket $[;]$ is called the Lie algebra of G and typically denoted $\mathfrak{g}$.

Special notations: $\mathfrak{gl}(n, \mathbb{R})$ is the Lie algebra of $\mathrm{GL}(n, \mathbb{R})$, $\mathfrak{o}(n)$ for $O(n)$, etc.

The general definition of a Lie algebra is a finite dimensional vector space together with a bracket on it that is alternating (so $[X, X] = 0$) and satisfies the Jacobi identity. When $[X, Y] = 0$ for all X, Y, we call the Lie algebra commutative.

Examples: The Lie algebra of

(1) $\mathbb{R}^n$ is $\mathbb{R}^n$ with trivial bracket,
(2) $\mathbb{S}^1$ is $\mathbb{R}$ with trivial bracket,
(3) $\mathrm{GL}(n, \mathbb{R})$ is $\mathbb{R}^{n^2}$ ($n \times n$ matrices) with bracket $[A, B] = AB - BA$.

Let's do that last computation. Since $\mathrm{GL}(n, \mathbb{R})$ is an open subset of the $n \times n$ matrices, its tangent space at I can be identified with the $n \times n$ matrices. Use x^i_j as coordinates on $\mathbb{R}^{n^2}$, so E^j_i denotes the matrix with 1 in the i-th row and j-th column and zeroes elsewhere. We then identify a matrix $M = (M^i_j)$ with the vector

$$M_I = \sum_{i,j} M^i_j E^j_i \in T_I \,\mathrm{GL}(n, \mathbb{R}).$$

(Here we are using e^j_i in place of $\frac{\partial}{\partial x^i_j}$ for ease of formula-reading.) Let $\tilde{M}$ denote the left invariant vector field with value M_I at I. For each $A \in \mathrm{GL}(n, \mathbb{R})$, the (k, ℓ) entry of $\tilde{M}$ at A is

$$\tilde{M}_A(x^k_\ell) = M_I(x^k_\ell \circ L_A),$$

where $x^k_\ell \circ L_A : \mathrm{GL}(n, \mathbb{R}) \to \mathbb{R}$ is the function mapping B to the (k, ℓ)-entry in AB, i.e.,

$$(x^k_\ell \circ L_A)(B) = \sum_{j=1}^n A^k_j B^j_\ell,$$

so that

$$\begin{aligned}
\frac{\partial}{\partial x^i_j}|_I (x^k_\ell \circ L_A) &= \lim_{h \to 0} \frac{(x^k_\ell)(I + hE^j_i) - (x^k_\ell \circ L_A)(I)}{h} \\
&= \lim_{h \to 0} \frac{1}{h} \left(\sum_{\alpha=1}^n \left(A^k_\alpha I^\alpha_\ell + h A^k_\alpha (E^j_i)^\alpha_\ell \right) - A^k_\ell \right) \\
&= \sum_{\alpha=1}^n A^k_\alpha (E^j_i)^\alpha_\ell \\
&= A^k_i (E^j_i)^i_\ell \begin{cases} A^k_i & j = \ell \\ 0 & j \neq \ell \end{cases} = A^k_i \delta^j_\ell
\end{aligned}$$

so that

$$\begin{aligned}\tilde{M}_A(x^k_\ell) &= M_I\left(x^k_\ell \circ L_A\right)\\ &= \sum_{i=1}^n\sum_{j=1}^n M^i_j A^k_i \delta^j_\ell\\ &= \sum_i M^i_\ell A^k_i = (AM)^k_\ell.\end{aligned}$$

In particular,

$$\begin{aligned}\frac{\partial}{\partial x^i_j}|_I(\tilde{M}(x^k_\ell)) &= \lim_{h\to 0}\frac{1}{h}\left(\tilde{M}(x^k_\ell)(I+hE^j_i) - \tilde{M}(x^k_\ell)(I)\right)\\ &= \left(E^j_i M\right)^k_\ell = \begin{cases} M_{j\ell} & k=i\\ 0 & k\neq i\end{cases}\end{aligned}$$

so that if N_I is another $n\times n$ matrix, we have that

$$N_I(\tilde{M}(x^k_\ell)) = (NM)^k_\ell,$$

and thus $[\tilde{M},\tilde{N}]_I = MN - NM$, i.e., $[M,N] = MN-NM$ in the sense of matrices (and so automatically satisfies the Jacobi identity).

11.2. **The exponential map.** Let $X\in\mathfrak{g}$ and let $\widetilde{X}$ denote the corresponding left invariant vector field on G. By our work earlier on differential equations, there is a unique one-parameter family φ_t of diffeomorphisms (that are also smooth in t) so that $\varphi_t = \mathrm{Id}$ and $\frac{d}{dt}\varphi_t(p) = \widetilde{X}_{\varphi_t(p)}$.

Proposition 107. *The map $\gamma:\mathbb{R}\to G$ given by $\gamma(t) = \varphi_t(e)$ is a homomorphism.*

Proof. We aim to show that $\gamma(s+t) = \gamma(s)\gamma(t)$ for all s and t. This follows from the observation that

$$\frac{d}{dt}(\gamma(t)) = \widetilde{X}_{\gamma(t)} = (L_{\gamma(t)})_* X.$$

Indeed, letting $\alpha_1(t) = \gamma(s+t)$ and $\alpha_2(t) = \gamma(s)\gamma(t)$, we see that α_1 satisfies the differential equation

$$\begin{aligned}\frac{d}{dt}\alpha_1(t) &= (L_{\alpha_1(t)})_* X,\\ \alpha_1(0) &= \gamma(s).\end{aligned}$$

On the other hand, we compute

$$\begin{aligned}\frac{d}{dt}\alpha_2(t) &= \frac{d}{dt}\left(L_{\gamma(s)}(\gamma(t))\right)\\ &= (L_{\gamma(s)})_*(\frac{d}{dt}\gamma(t)) = (L_{\gamma(s)})_*(L_{\gamma(t)})_* X.\end{aligned}$$

As $F_*\circ G_* = (F\circ G)_*$, we have that $(L_{\gamma(s)})_*(L_{\gamma(t)})_* = (L_{\gamma(s)}\circ L_{\gamma(t)})_* = (L_{\gamma(s)\gamma(t)})_*$ and thus α_2 satisfies

$$\begin{aligned}\frac{d}{dt}\alpha_2(t) &= (L_{\alpha_2(t)})_* X,\\ \alpha_2(0) &= \gamma(s).\end{aligned}$$

As the two curves satisfy the same initial value problem, the uniqueness theorem implies that they agree, i.e., $\gamma(s+t) = \gamma(s)\gamma(t)$. $\square$

Corollary 108. *For any Lie group G and $X \in \mathfrak{g}$, there is a unique smooth homomorphism $\phi : \mathbb{R} \to G$ with $\frac{d\phi}{dt}|_{t=0} = X$.*

Assuming $X \neq 0$, the homomorphism defined above is injective; its image is called a 1-parameter subgroup of G.

We use this homomorphism to define the exponential map. Indeed, for any Lie group G, we define the exponential map $\exp : \mathfrak{g} \to G$. Suppose $X \in \mathfrak{g}$ and let $\phi : \mathbb{R} \to G$ denote the homomorphism above; we define $\exp(X) = \phi(1)$.

Theorem 109. *The exponential map $\exp : \mathfrak{g} \to G$ satisfies*

(1) $\exp((t_1 + t_2)X) = \exp(t_1 X)\exp(t_2 X)$,
(2) $\exp(-tX) = \exp(tX)^{-1}$,
(3) 0 *is a regular point of* $\exp : \mathfrak{g} \to G$, *so* $\exp$ *is a diffeomorphism from a neighborhood of* 0 *in* $\mathfrak{g}$ *to a neighborhood of* e *in* G, *and*
(4) if $\psi : G \to H$ *is a smooth homomorphism, then* $\exp_H \circ \psi_* = \psi \circ \exp_G$.

Proof. The first two follow immediately by uniqueness. We turn our attention to the third part.

Using the identification $T_0\mathfrak{g} \cong \mathfrak{g}$, consider the curve $c(t) = tv \in T_eG$; its tangent vector at $t = 0$ is v. We then have, essentially by definition, that

$$(\exp_*)_0(v) = \frac{d}{dt}|_{t=0}(\exp(tv)) = v,$$

so that $(\exp_*)_0 = \mathrm{Id}_{\mathfrak{g}}$, which is an isomorphism and so 0 is a regular point.

Now, let $\psi : G \to H$ be a smooth homomorphism and $X \in \mathfrak{g} = T_eG$, and let $\phi : \mathbb{R} \to G$ denote the smooth homomorphism with $\frac{d\phi}{dt}|_{t=0} = X$ so that $\phi(t) = \exp(tX)$. Then $\psi \circ \phi : \mathbb{R} \to H$ is a smooth homomorphism with

$$\frac{d}{dt}|_{t=0}(\psi \circ \phi) = \psi_* X,$$

and thus $\exp_H(\psi_* X) = \psi \circ \phi(1) = \psi \circ \exp_G X$. $\square$

Corollary 110. *If $\phi : G \to H$ is a smooth homomorphism that is also injective, then ϕ is an immersion.*

Proof. If not, then there is some $p \in G$ and a non-zero $X \in \mathfrak{g}$ with $(\phi_*)_p(\widetilde{X}_p) = 0$, so then $(\phi_*)_e X = 0$ and thus

$$e_H = \exp_H(\psi_*(tX)) = \phi(\exp_G(tX)),$$

but ϕ is injective, so $\exp_G(tX) = e_G$ for all t and thus $X = 0$, a contradiction. $\square$

Corollary 111. *Every continuous homomorphism $\mathbb{R} \to G$ is smooth.*

Proof. Let $\phi : \mathbb{R} \to G$ be a continuous homomorphism. Because it is a homomorphism, we can compose with translations and thus it suffices to check that it is smooth in a neighborhood of 0.

Let $U \subset \mathfrak{g}$ be a star-shaped neighborhood of 0 on which exp is one-to-one. For any $a \in \exp(\frac{1}{2}U)$, write $a = \exp(\frac{1}{2}X)$ for $X \in U$, so that

$$a = \exp\left(\frac{1}{4}X + \frac{1}{4}X\right) = \exp\left(\frac{1}{4}X\right)^2.$$

In particular, each $a \in \exp(\frac{1}{2}U)$ has a square root in $\exp(\frac{1}{2}U)$.

In fact, the square root is unique in this set: If $a = b^2$ for $b \in \exp(\frac{1}{2}U)$, then we write $b = \exp(\frac{1}{2}Y)$ and then

$$\exp\left(\frac{1}{2}X\right) = \exp\left(\frac{1}{2}Y\right)^2 = \exp(Y),$$

so that $Y = \frac{X}{2}$ and $Y/2 = X/4$ as exp is injective in U.

We now pick $\epsilon > 0$ so that $\phi(t) \in \exp(\frac{1}{2}U)$ for $|t| \leq \epsilon$. This can be done because ϕ is continuous and $\phi(0) = e$. Let $X \in \frac{1}{2}U$ be such that $\phi(\epsilon) = \exp(X)$. We now observe that

$$\phi\left(\frac{\epsilon}{2}\right)^2 = \phi(\epsilon) = \exp\left(\frac{X}{2}\right)^2,$$

so by the uniqueness of square roots in this set we must have $\phi(\epsilon/2) = \exp(X/2)$. Similarly, by repeating this argument and then using that ϕ is a homorphism, we obtain

$$\phi\left(\frac{m\epsilon}{2^k}\right) = \exp\left(\frac{mX}{2^k}\right)$$

for all $k \geq 0$ and $m \in \mathbb{Z}$, $|m| \leq 2^k$. By continuity, we therefore have $\phi(s\epsilon) = \exp(sX)$ for all $s \in [-1, 1]$. □

Corollary 112. *If G, H are Lie groups, any continuous homomorphism $\phi : G \to H$ is smooth.*

Proof. Again it suffices to check in a neighborhood of the identity. Choose a basis $X_1, \ldots, X_n$ for $\mathfrak{g}$. $t \mapsto \phi(\exp_G(tX_i))$ is a continuous homomorphism $\mathbb{R} \to H$ and so there are $Y_i \in \mathfrak{h}$ with $\phi(\exp_G(tX_i)) = \exp_H(tY_i)$. Now

$$\psi : \mathbb{R}^n \to G, \quad \psi(t^1, \ldots, t^n) = \exp(t^1 X_1) \ldots \exp(t^n X_n)$$

is a smooth map with $\psi_*(\frac{\partial}{\partial t^i}) = X_i$, so ψ is a local diffeomorphism. Note that

$$\phi \circ \psi(t^1, \ldots, t^n) = \exp_H(t^1 Y_1) \ldots \exp_H(t^n Y_n),$$

which is smooth by the previous corollary (and the smoothness of the multiplication map on H. Writing $\phi = (\phi \circ \psi) \circ \psi^{-1}$ we see that ϕ is smooth. □

Corollary 113. *If G and G' are Lie groups that are isomorphic as topological groups then they are also isomorphic as Lie groups.*

The following theorem will also be useful in our discussion of subgroups.

Theorem 114. *If G is a Lie group and $X, Y \in \mathfrak{g}$, then*

(1) $\exp(tX)\exp(tY) = \exp\left(t(X+Y) + \frac{t^2}{2}[X,Y] + O(t^3)\right)$,

(2) $\exp(-tX)\exp(-tY)\exp(tX)\exp(tY) = \exp\left(\frac{t^2}{2}[X,Y] + O(t^3)\right)$, *and*

Here $O(t^3)$ denotes a function $c(t) : \mathbb{R} \to \mathfrak{g}$ with the property that $c(t)/t^3$ is bounded as $t \to 0$.

Proof. As usual, let $\widetilde{X}$ and $\widetilde{Y}$ denote the left invariant vector fields extending X and Y. Suppose $f : G \to \mathbb{R}$ is smooth.

We first observe that

$$(\widetilde{X}f)(a) = \widetilde{X}_a f = X(f \circ L_a) = \frac{d}{ds}|_{s=0} f(a \exp(sX)),$$

with a similar expression holding for $\widetilde{Y}f$.

Now, fixing s, let

$$\phi(t) = f\left(\exp(sX)\exp(tY)\right),$$

so that

$$\begin{aligned}\phi'(t) &= \frac{d}{dt} f\left(\exp(sX)\exp(tY)\right)\\ &= \frac{d}{du}|_{u=0} f\left(\exp(sX)\exp(tY)\exp(uY)\right)\\ &= (\widetilde{Y}f)\left(\exp(sX)\exp(tY)\right).\end{aligned}$$

Applying again to $\widetilde{Y}f$ yields that

$$\phi''(t) = [\widetilde{Y}(\widetilde{Y}f)]\left(\exp(sX)\exp(tY)\right).$$

Taylor's theorem then tells us that

$$\phi(t) = \phi(0) + \phi'(0)t + \frac{1}{2}\phi''(0)t^2 + O(t^3).$$

In particular, if $f(e) = 0$, then

$$f\left(\exp(sX)\exp(tY)\right) = f(\exp(sX)) + t(\widetilde{Y}f)(\exp(sX)) + \frac{t^2}{2}(\widetilde{Y}(\widetilde{Y}f))(\exp(sX)) + O(t^3).$$

A similar story holds for X and some smooth F, so that

$$\begin{aligned}\frac{d}{ds}F(\exp sX) &= (\widetilde{X}F)(\exp sX),\\ \frac{d^2}{ds^2}F(\exp sX) &= [\widetilde{X}(\widetilde{X}f)](\exp sX),\\ F(\exp sX) &= F(e) + s(\widetilde{X}F)(e) + \frac{s^2}{2}[\widetilde{X}(\widetilde{X}f)](e) + O(t^3).\end{aligned}$$

Substituting in, we get, if $f(e) = 0$,

$$\begin{aligned}f\left(\exp(sX)\exp(tY)\right) = {} & s(\widetilde{X}f)(e) + \frac{s^2}{2}(\widetilde{X}(\widetilde{X}f))(e) + O(s^3)\\ & + t(\widetilde{Y}f)(e) + st(\widetilde{X}(\widetilde{Y}f))(e) + O(s^2 t)\\ & + \frac{t^2}{2}(\widetilde{Y}(\widetilde{Y}f))(e) + O(st^2) + O(t^3).\end{aligned}$$

We now take $s = t$; for small t write $\exp(tX)\exp(tY) = \exp(Z(t))$, where $Z : (-\epsilon, \epsilon) \to \mathfrak{g}$ is smooth. The first few terms of its Taylor expansion are

$$Z(t) = 0 + tZ_1 + \frac{t^2}{2}Z_2 + O(t^3).$$

If f is smooth and $f(e) = 0$, then $f(\exp(A(t) + O(t^3))) = f(A(t)) + O(t^3)$, so

$$f(\exp Z(t)) = t(\tilde{Z}_1 f)(e) + \frac{t^2}{2}(\tilde{Z}_2 f)(e) + \frac{t^2}{2}(\tilde{Z}_1(\tilde{Z}_1 f))(e) + O(t^3).$$

After taking the f to be coordinate functions, we match up the power series to get that

$$\widetilde{X} + \widetilde{Y} = \tilde{Z}_1,$$
$$\frac{\tilde{Z}_1\tilde{Z}_1}{2} + \tilde{Z}_2 = \frac{1}{2}\widetilde{X}\widetilde{X} + \frac{1}{2}\widetilde{Y}\widetilde{Y} + \widetilde{X}\widetilde{Y},$$

so that $Z_1 = X + Y$ and $Z_2 = \frac{1}{2}[X, Y]$, proving the first two results. The □

11.3. **Subgroups, etc.**

Definition 115. A Lie subgroup H of a Lie group G is a subgroup $H \subset G$ that has a smooth structure making the inclusion $i : H \to G$ an immersion.

(Might just be immersed, not embedded – image of lines with irrational slopes are fine Lie subgroups of $\mathbb{S}^1 \times \mathbb{S}^1$ but are dense.

Definition 116. A Lie subalgebra of $\mathfrak{g}$ is a subspace of $\mathfrak{g}$ that is closed under the Lie bracket.

It's not hard to show that if $H \subset G$ is a Lie subgroup , then the Lie algebra of H is a Lie subalgebra of $\mathfrak{g}$. A converse is also true:

Theorem 117. *If G is a Lie group and $\mathfrak{h}$ is a Lie subalgebra of the Lie algebra of G, then there is a unique connected Lie subgroup H of G with Lie algebra $\mathfrak{h}$.*

Proof. For each $X \in \mathfrak{h}$, let $\widetilde{X}$ denote the corresponding left invariant vector field with value X at e. At each $a \in G$, let $\Delta_a \subset T_aG$ be the subspace spanned by $\widetilde{X}_a$ for all $X \in \mathfrak{h}$. As $\mathfrak{h}$ is a Lie subalgebra, Δ is closed under the Lie bracket and thus is an integrable distribution. Let H be the maximal integral submanifold of Δ containing e.

Left-invariance implies that if $b \in G$, $(L_b)_*(\Delta_a) = \Delta_{ba}$, so $(L_b)_*$ leaves Δ invariant and thus L_b only permutes the integral submanifolds. In particular, if $b \in H$, then $L_{b^{-1}}(H) = H$, so H is a subgroup. □

Suppose now that G and H are Lie groups and $\phi : G \to H$ is a smooth homomorphism, so that $\phi_* : \mathfrak{g} \to \mathfrak{h}$. For any $a \in G$, $\phi \circ L_a = L_{\phi(a)} \circ \phi$ because ϕ is a homomorphism. In particular, for any $X \in \mathfrak{g}$ and $Y = \phi_* X$, if $\widetilde{X}$ is the corresponding left invariant vector field on G and $\widetilde{Y}$ the analogous object on H, we have

$$(\phi_*)_a \widetilde{X}_a = (\phi_*)_a ((L_a)_* X) = (L_{\phi(a)})_*(\phi_* X) = (L_{\phi(a)})_* Y = \widetilde{Y}_{\phi(a)},$$

so that $\widetilde{X}$ and $\widetilde{Y}$ are ϕ-related and thus

$$\phi_*[X, Y] = [\phi_* X, \phi_* Y],$$

i.e., ϕ_* is a Lie algebra homomorphism.

The following theorem tells us that Lie algebra homomorphisms lift to "local homomorphisms".

Theorem 118. *Suppose G and H are Lie groups and $\Phi : \mathfrak{g} \to \mathfrak{h}$ is a Lie algebra homomorphism. There is a neighborhood U of e in G and a smooth map $\phi : U \to H$ so that $\phi(ab) = \phi(a)\phi(b)$ whenever $a, b, ab \in U$ and so that $(\phi_*)_e = \Phi$. Moreover, if there are smooth homomorphisms $\phi, \psi : G \to H$ with $\phi_* = \psi_* = \Phi$ and G is connected, then $\phi = \psi$.*

Note that the theorem does not tell us that such a homomorphism exists; if G is simply connected, one can show that ϕ extends to a global homomorphism.

Proof. Let $\mathfrak{k} \subset \mathfrak{g} \times \mathfrak{h}$ be

$$\mathfrak{k} = \{(X, \Phi(X)) : X \in \mathfrak{g}\}.$$

As Φ is a homomorphism, $\mathfrak{k}$ is a Lie subalgebra of $\mathfrak{g} \times \mathfrak{h}$, which is the Lie algebra of $G \times H$, so there is a unique connected Lie subgroup K of $G \times H$ with Lie algebra $\mathfrak{k}$.

Now let $\pi_1 : G \times H \to G$ denote the projection (a smooth homomorphism!) and let $\theta = \pi_1|_K$, so that $\theta : K \to G$ is a smooth homomorphism.

For $X \in \mathfrak{g}$, $\theta_*(X, \Phi(X)) = X$, so $\theta_* : T_{(e,e)}K \to T_eG$ is an isomorphism. The inverse function theorem gives an open neighborhood and a local inverse which must locally be a homomorphism, so we take $\phi = \pi_2 \circ \theta^{-1}$, and then

$$\phi_*(X) = (\pi_2)_*((\theta^{-1})_*X) = (\pi_2)_*(X, \Phi(X)) = \Phi(X),$$

as desired. The uniqueness statement is built into the uniqueness of K. $\square$

Corollary 119. *If G and H are Lie groups with isomorphic Lie algebras then G and H are locally isomorphic.*

Corollary 120. *The Lie bracket of $\mathfrak{g}$ is trivial (i.e., $[X, Y] = 0$ for all $X, Y \in \mathfrak{g}$) if and only if the connected component of G containing the identity is abelian.*

Proof. If G is abelian, then $g \mapsto g^{-1}$ is a smooth homomorphism $i : G \to G$. The derivative of this inversion map $i_* : \mathfrak{g} \to \mathfrak{g}$ is given by $i_*X = -X$. As the derivative of a Lie group homomorphism descends to a Lie algebra homomorphism, we must have

$$-[X, Y] = i_*[X, Y] = [i_*X, i_*Y] = [-X, -Y] = [X, Y],$$

so that $[X, Y] = 0$.

Now suppose $[X, Y] = 0$ for all X and Y, i.e., $\mathfrak{g}$ is isomorphic to the Lie algebra of $\mathbb{R}^n$ for some n. By the previous corollary, G must therefore be locally isomorphic to $\mathbb{R}^n$, and thus the connected component containing the identity must be abelian. $\square$

The hard work at the end of the last subsection (which we didn't/won't do in class) helps prove the following theorem (also omitted):

Theorem 121. *If G is a Lie group and $H \subset G$ is a closed subset that is also a subgroup, then H is a Lie subgroup of G.*

12. Curves and surfaces

12.1. **Curves in $\mathbb{R}^2$.** Throughout this section, we assume that $\gamma : [a, b] \to \mathbb{R}^2$ (or $\mathbb{R}^3$) is a smooth curve so that γ is also an immersion (i.e., $\gamma'(t) \neq 0$ for all $t \in [a, b]$).

Definition 122. The length of γ is $\int_a^b |\gamma'(t)| \, dt$.

Note that γ is independent of parametrization! We say that γ is parametrized by arc length if $|\gamma'(s)| = 1$ for all s.

Proposition 123. *Any such curve can be preparametrized by arc length.*

Proof. Suppose $\tilde{\gamma}(s) = \gamma \circ \sigma$, so

$$\tilde{\gamma}'(s) = \frac{d\gamma}{dt}(\sigma(s))\frac{d\sigma}{ds}.$$

To ensure that $\tilde{\gamma}$ is parametrized by arc length, we want

$$\left|\frac{d\sigma}{ds}\right| = \frac{1}{|\gamma'(\sigma(s))|}.$$

As $\gamma' \neq 0$, this differential equation can be solved. □

Now suppose $\gamma : [0, L] \to \mathbb{R}^2$ is parametrized by arc length. Let $t(s) = \gamma'(s)$ be the (unit-length!) tangent vector and $n(s)$ denote the unit vector in $\mathbb{R}^2$ so that $t(s) \cdot n(s) = 0$ and so that (t, n) gives the standard orientation of $\mathbb{R}^2$. In terms of the components (t_1, t_2) of t, this requirement implies that $n = (-t_2, t_1)$; n is of course also smooth.

Since $|\gamma'(s)| = 1$, we must have

$$0 = \frac{d}{ds}|\gamma'(s)|^2 = 2\langle\gamma'(s), \gamma''(s)\rangle,$$

so $t'(s) = \gamma''(s)$ is normal to $t(s) = \gamma'(s)$ and thus there is a function $\kappa(s)$ so that $t'(s) = \kappa(s)n(s)$. Observe that our explicit characterization of $n(s)$ in terms of t therefore implies that $n'(s) = -\kappa(s)t(s)$.

Definition 124. $\kappa(s)$ is the curvature of the plane curve γ.

As an example, you should check that the curvature of a circle of radius R is $1/R$.

The following proposition follows from the homotopy lifting property enjoyed by $\mathbb{R}$ as the universal cover of $\mathbb{S}^1$; you either have seen it in a topology class or will see it in a topology class. Let $\mu : \mathbb{R} \to \mathbb{S}^1$ be given by $\mu(\theta) = (\cos\theta, \sin\theta)$.

Proposition 125. *If $\gamma : [a, b] \to \mathbb{S}^1$ is continuous, then there is a continuous function $f : [a, b] \to \mathbb{R}$ with $\mu \circ f = \gamma$. If f and $\bar{f}$ are two such functions, then $f - \bar{f} = 2\pi k$ for some constant $k \in \mathbb{Z}$.*

Now suppose γ is a curve parametrized by arc length, then $t = \gamma'(s)$ is a map $[0, L] \to \mathbb{S}^1$. Letting f denote the lift of t, we then have

$$\gamma'(s) = t(s) = \mu(f(s)) = (\cos f(s), \sin f(s)),$$

so that

$$\gamma''(s) = (-f'(s)\sin f(s), f'(s)\cos f(s)),$$

i.e., $\kappa(s) = f'(s)$. The total curvature of γ is therefore $\int_0^L \kappa(s)\, ds = f(b) - f(a)$.

Suppose now that γ is a closed curve parametrized by arc length, so $t : [0, L] \to \mathbb{S}^1$ is also a closed curve. With f as above for t, note that because $t(L) = t(0)$, we must have $f(L) = f(0) + 2\pi k$ for some integer k. We say that k is the degree of t and denote it $\deg t$. We also say that it is the rotation number of t.

Proposition 126. *If $\gamma : [0, L] \to \mathbb{R}^2$ is a closed curve, then the total curvature of γ is $2\pi \deg t$.*

Proof. This is almost by definition (as we've defined things in this course):

$$\deg t = \frac{1}{2\pi}(f(L) - f(0)) = \frac{1}{2\pi}\int_0^L f'(s)\, ds = \frac{1}{2\pi}\int_0^L \kappa(s)\, ds.$$

□

Here's an example of more stuff you can do in two dimensions:

Theorem 127. *A vertex of a smooth curve in $\mathbb{R}^2$ parametrized by arc length is one where the curvature has a critical point, i.e., $\kappa'(s) = 0$. Every simple closed curve has at least four vertices.*

Proof. It's pretty easy to find at least two: look at the maximum and minimum of $\kappa(s)$. Now change the coordinate stystem so that the x-axis passes through these two, dividing the curve into an upper and lower half. We calculate:

$$\begin{aligned}\int_0^L \kappa'(s)\gamma(s)\,ds &= -\int_0^L \kappa(s)t(s)\,ds \\ &= \int_0^L n'(s)\,ds = n(L) - n(0) = 0.\end{aligned}$$

Taking the inner product with e_2 yields that

$$\int_0^L \kappa'(s)\langle\gamma(s), e_2\rangle\,ds = 0.$$

If there are no more vertices, then $\kappa'(s) > 0$ on one of the two halves and < 0 on the other half, so that $\kappa'(s)\langle\gamma(s), e_2\rangle$ has the same sign on both halves, a contradiction. There must then be at least three vertices; if there are only three then $\kappa'(s)$ must still have the same sign on two "halves"; repeating the above argument in a new coordinate system yields a contradiction so in fact there must be at least four vertices. □

12.2. **Curves in $\mathbb{R}^3$.** Now suppose $\gamma : [0, L] \to \mathbb{R}^3$ is parametrized by arc length. We define:

- the unit tangent $t(s) = \gamma'(s)$,
- the curvature vector $\gamma''(s) = t'(s)$,
- the curvature $\kappa(s) = |\gamma''(s)|$,
- the unit normal $n(s) = \frac{t'(s)}{|t'(s)|}$ (only when $\kappa \neq 0$),
- and the binormal $b(s) = t(s) \times n(s)$ (note 3-dimensions!)

Observe that when $\kappa \neq 0$, (t, n, b) is a positively oriented orthonormal basis of $\mathbb{R}^3$.

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 128. 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 129. 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 130. 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 131. *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 132. *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 $\mathcal{I}_p$. □

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

Lemma 133. *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. □

Proof of Proposition 131. 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 133. □

From the proof of Proposition 131, 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 $F^* f = f \circ F \in C^\infty(U)$.

Appendix B. Smooth dependence on parameters

The aim of this section is to prove Theorem 44. In fact, we prove a somewhat stronger theorem that implies Theorem 44.

Consider first the initial value problem

$$\frac{dy}{dt} = f(t, y),$$
$$y(t_0) = a,$$

for $f : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$, as well as a similar problem depending on parameters

$$\frac{dy}{dt} = f(t, y, \mu),$$
$$y(t_0) = a,$$

where $\mu \in \mathbb{R}^k$.

We are interested in how each of the problems above depends on parameters; in the first problem, we are interested in the dependence on initial value a, while in the second problem we are interested in the dependence on a as well as μ. Note that by taking $\tilde{f}(t, y, a) = f(t, y + a)$, we can think of the initial value problem as the problem

$$\frac{dy}{dt} = \tilde{f}(t, y, a),$$
$$y(t_0) = 0,$$

to a problem with fixed initial value depending on parameters. If y solves this problem, then $y + a$ will solve the original initial value problem. If, instead, one knows how to solve the initial value problem, one can augment the differential equation to include $d\mu/dt = 0$ and $\mu(t_0) = 0$ to solve the problem with parameters.

We'll focus on solving the problem with parameters. We'll assume f is smooth in all of its arguments. Note that our proof of Picard iteration easily adapts (the smoothness statement follows from repeated differentiation) to show the following:

Theorem 134. *Suppose $I \subset \mathbb{R}$, $0 \in U \subset \mathbb{R}^n$, and $V \subset \mathbb{R}^k$ are open and that $f : I \times U \times V \to \mathbb{R}^n$ is smooth. If $(t_0, 0, \mu_0) \in I \times U \times V$ then there are $b > 0$, and $\epsilon > 0$ so that for all $\mu \in \overline{B_\epsilon(\mu_0)}$ there is a smooth function $\gamma_\mu : (t_0 - b, t_0 + b) \to \mathbb{R}^n$ satisfying*

$$\frac{d\gamma_\mu}{dt} = f(t, \gamma_\mu(t), \mu),$$
$$\gamma_\mu(t_0) = 0.$$

Our aim is to prove the following theorem:

Theorem 135. *Suppose $I \subset \mathbb{R}$, $0 \in U \subset \mathbb{R}^n$, and $V \subset \mathbb{R}^k$ are open and that $f : I \times U \times V \to \mathbb{R}^n$ is smooth. Let $\gamma_\mu(t)$ denote the solution of*

$$\frac{d\gamma_\mu}{dt} = f(t, \gamma_\mu(t), \mu)$$
$$\gamma_\mu(t_0) = 0.$$

If $(t_0, x_0, \mu_0) \in I \times U \times V$ then there are $b > 0$ and $\epsilon > 0$ so that $\gamma_\mu(t)$ is smooth (as a function of (t, μ) on $(t_0 - b, t_0 + b) \times B_\epsilon(\mu_0)$.

We'll prove this theorem in three steps. First we'll show that the map $(t, \mu) \mapsto \gamma_{x,\mu}(t)$ is continuous, then show that it is differentiable and finally (and most easily) show that it is smooth. In what follows it helps to have a short definition. For an set $U \subset \mathbb{R}^k$, we say that a function $f : U \to \mathbb{R}^n$ is Lipschitz with Lipshitz constant L if for all $x, y \in U$, $|f(x) - f(y)| \le L\,|x - y|$.

The following theorem provides a condition guaranteeing that the solution depends continuously on the parameter on finite time intervals.

Theorem 136. *Suppose, for some $b > 0$ and open sets $I = (t_0 - b, t_0 + b) \subset \mathbb{R}$, $U \subset \mathbb{R}^n$ and $V \subset \mathbb{R}^k$, we have*

$$f : I \times U \times V \to \mathbb{R}^n$$

is continuous. Suppose further that, for each $(t, \mu) \in I \times V$, $f(t, \cdot, \mu)$ is Lipschitz with Lipschitz constant L_1 and that for each $(t, y) \in I \times U$, $f(t, y, \cdot)$ is Lipschitz with Lipschitz constant L_2. If $y_i : I \to \mathbb{R}^n$ $(i = 1, 2)$ satisfy

$$\frac{dy_i}{dt} = f(t, y_i(t), \mu_i),$$
$$y_i(t_0) = 0,$$

then for all $t \in I$,

$$|y_1(t) - y_2(t)| \le \frac{L_2}{L_1}\,|\mu_1 - \mu_2|\left(e^{L_1|t-t_0|} - 1\right).$$

A useful tool here is Gronwall's inequality:

Lemma 137 (Gronwall inequality)**.** *Suppose that $X(t)$ is a continuous, real-valued function on $[t_0, T]$. If there are constants $C, K > 0$ so that*

$$X(t) \le C + K\int_{t_0}^{t} X(s)\,ds$$

for every $t \in [t_0, T]$, then

$$X(t) \le Ce^{K(t-t_0)}$$

for all $t \in [t_0, T]$.

Proof. Define $v(t) = e^{-K(t-t_0)}\int_{t_0}^{t} X(s)\,ds$. Observe that

$$v'(t) = e^{-K(t-t_0)}\left(X(t) - K\int_{t_0}^{t} X(s)\,ds\right) \le Ce^{-K(t-t_0)}.$$

As $v(t_0) = 0$, we can integrate to see that, for $t \in [t_0, T]$,

$$v(t) \le C\int_{t_0}^{t} e^{-K(s-t_0)}\,ds = \frac{C}{K}\left(1 - e^{-K(t-t_0)}\right),$$

i.e.,

$$\int_{t_0}^{t} X(s)\,ds \le \frac{C}{K}\left(e^{K(t-t_0)} - 1\right),$$

so that

$$X(t) \le C + K\int_{t_0}^{t} X(s)\,ds \le C + Ce^{K(t-t_0)} - C = Ce^{K(t-t_0)}.$$

$\square$

Proof of Theorem 136. We'll consider only $t \geq t_0$ for simplicity; the reader should fill in the details for $t < t_0$. Observe that

$$\begin{aligned}
|y_1(t) - y_2(t)| &= \left| \int_{t_0}^{t} \left(f(s, y_1(s), \mu_1) - f(s, y_2(s), \mu_2) \right) \, ds \right| \\
&\leq \int_{t_0}^{t} |f(s, y_1(s), \mu_1) - f(s, y_2(s), \mu_2)| \, ds \\
&\leq \int_{t_0}^{t} \left(|f(s, y_1(s), \mu_1) - f(s, y_2(s), \mu_1)| + |f(s, y_2(s), \mu_1) - f(s, y_2(s), \mu_2)| \right) \, ds \\
&\leq \int_{t_0}^{t} \left(L_1 |y_1(s) - y_2(s)| + L_2 |\mu_1 - \mu_2| \right) \, ds.
\end{aligned}$$

We now let $X(t) = L_1 |y_1(s) - y_2(s)| + L_2 |\mu_1 - \mu_2|$, so that

$$X(t) \leq L_2 |\mu_1 - \mu_2| + L_1 \int X(s) \, ds.$$

By the Gronwall inequality we conclude that $X(t) \leq L_2 |\mu_1 - \mu_2| e^{L_1(t-t_0)}$. In particular,

$$|y_1(s) - y_2(s)| \leq \frac{L_2}{L_1} |\mu_1 - \mu_2| \left(e^{L_1(t-t_0)} - 1 \right),$$

as claimed. ☐

We now turn our attention to the problem of differentiable dependence. The formal idea is to differentiate the equation; the hard part is showing that in fact this is justified.

Theorem 138. *Suppose $f : I \times U \times V \to \mathbb{R}^n$ satisfies the hypotheses of Theorem 136. If f in addition is continuously differentiable with respect to y and μ, then the solution $y(t, \mu)$ of*

$$\begin{aligned}
\frac{dy}{dt} &= f(t, y(t, \mu), \mu), \\
y(t_0) &= 0,
\end{aligned}$$

is differentiable with respect to μ, and $z_j = \frac{\partial y}{\partial \mu^j}$ satisfies

$$\begin{aligned}
\frac{dz_j}{dt} &= \sum_{k=1}^{n} \frac{\partial f}{\partial y^k}(t, y(t, \mu), \mu) z_j^k + \frac{\partial f}{\partial \mu^j}(t, y(t, \mu), \mu), \\
z_j(t_0) &= 0.
\end{aligned}$$

Proof. Our plan is to show first that the equation for z has a solution and then to use Gronwall's inequality to show that the difference quotients for $\frac{\partial y}{\partial \mu_j}$ converge to z.

It is straightforward to see that the right side of the ODE for z is continuous in t and z and Lipshitz in z and so has a unique solution $z(t, \mu)$ for each μ. Now define w_j to be the difference quotient for y in the j-th direction (here e_j is the j-th standard basis vector in $\mathbb{R}^n$).

$$w_j(t, \mu, h) = \frac{y(t, \mu + he_j) - y(t, \mu)}{h}.$$

We fix j for now and drop the subscripts on w and z. We want to show that $w(t, \mu, h) \to z(t, \mu)$ uniformly on compact sets as $h \to 0$. Let $v(t, \mu, h) = w(t, \mu, h) - z(t, \mu)$ denote their difference. We compute

$$\frac{dv}{dt} = \frac{dw}{dt} - \sum_k \frac{\partial f}{\partial y^k}(t, y(t,\mu), \mu) z^k(t,\mu) - \frac{\partial f}{\partial \mu_j}(t, y(t,\mu),\mu).$$

Because y satisfies an ODE, the first term in the above is given by

$$\begin{aligned}\frac{dw}{dt}(t,\mu,h) &= \frac{f(t, y(t,\mu + he_j), \mu + he_j) - f(t, y(t,\mu),\mu)}{h} \\ &= \frac{f(t, y(t,\mu + he_j), \mu + he_j) - f(t, y(t,\mu),\mu + he_j)}{h} + \frac{f(t, y(t,\mu), \mu + he_j) - f(t, y(t,\mu),\mu)}{h}.\end{aligned}$$

Appealing to the first-order version of Taylor's theorem (i.e., the fundamental theorem of calculus), we have, because f is C^1 in y and μ,

$$\begin{aligned}\frac{dw}{dt}(t,\mu,h) &= \sum_{k=1}^n \left(\int_0^1 \frac{\partial f}{\partial y^k}(t, sy(t,\mu + he_j) + (1-s)y(t,\mu), \mu + he_j)\, ds\right) w^k(t,\mu,h) \\ &\quad + \int_0^1 \frac{\partial f}{\partial \mu^j}(t, y(t,\mu), \mu + she_j)\, ds.\end{aligned}$$

The differential equation for v can therefore be written

$$\begin{aligned}\frac{dv}{dt} &= \sum_{k=1}^n \left(\int_0^1 \frac{\partial f}{\partial y^k}(t, sy(t,\mu + he_j) + (1-s)y(t,\mu), \mu + he_j)\, ds\right) v^k(t,\mu,h) \\ &\quad + \sum_{k=1}^n \int_0^1 \left(\frac{\partial f}{\partial y^k}(t, sy(t,\mu + he_j) + (1-s)y(t,\mu), \mu + he_j) - \frac{\partial f}{\partial y^k}(t, y(t,\mu),\mu)\right) ds\, z^k(t,\mu) \\ &\quad + \int_0^1 \left(\frac{\partial f}{\partial \mu^j}(t, y(t,\mu), \mu + she_j) - \frac{\partial f}{\partial \mu^j}(t, y(t,\mu),\mu)\right) ds, \\ v(t_0,\mu,h) &= 0.\end{aligned}$$

We now regard the differential equation for v as a parameter-dependent family of differential equations depending on a parameter h. Theorem 136 implies that the right hand side of this differential equation is Lipschitz in h; it is clearly Lipschitz in v. We may thus conclude that $v(t,\mu,h)$ converges uniformly on compact sets in t to $v(t,\mu,0)$. When $h = 0$, the right hand side vanishes and $v(t,\mu,0) = 0$ by uniqueness, hence $w_j \to z_j$, i.e., $\frac{\partial y}{\partial \mu_j}$ exists and is given by z_j.

To see that in fact $\frac{\partial y}{\partial \mu_j}$ is continuous, we appeal again to Theorem 136; the right hand side of the differential equation satisfied by z_j is Lipschitz in z and μ and so z must depend continuously on μ. □

We finally reach the conclusion of this section:

Corollary 139. *If, in addition, f is C^∞, then the solution $y(t,\mu)$ is also C^∞.*

Proof. Inductively apply Theorem 138 to conclude that each $\frac{\partial^{|\alpha|} y}{\partial \mu^\alpha}$ is continuous and so y is smooth. □