

Introduction to Differential Manifolds

Clayton Shonkwiler

Contents

1	Manifolds and Vector Fields	3
1.1	Manifolds and Maps	3
1.2	Tangent Vectors	7
1.3	The Differential	11
1.4	Immersions and Embeddings	16
1.5	Vector Fields	20
1.5.1	The matrix exponential	22
1.6	The Lie Bracket on Vector Fields	24
1.7	Integral Curves and Lie Derivatives	30
1.8	Extended Example on S^3	35
2	Differential Forms	39
2.1	Integration, Forms, and an Informal Definition	39
2.2	Differential Forms and Vector Calculus	43
2.3	Tensor Algebras and Tensor Fields	46
2.4	Exterior Algebras and Differential Forms	51
2.4.1	Differential Forms in Local Coordinates	55
2.5	The Exterior Derivative	56
2.6	Pullbacks	59
2.7	Lie Derivatives and Cartan's Magic Formula	63
2.8	De Rham Cohomology	70
2.9	Integration on Manifolds	71
2.9.1	Digression on Differentials, Pullbacks, and Dual Maps	71
2.9.2	How Differential Forms Transform	71
2.9.3	Definition of the Integral	73
2.10	Integration on Submanifolds	76
2.11	Stokes' Theorem	79

3 Lie Groups and Homogeneous Spaces	82
3.1 Some Motivation from de Rham Cohomology	82
3.2 Lie groups	83
3.3 Lie Algebras	85
3.4 Matrix Groups	88

Chapter 1

Manifolds and Vector Fields

1.1 Manifolds and Maps

The title of this course is “Introduction to Differential Manifolds,” which suggests that these *differential manifolds* (or sometimes *differentiable manifolds*), whatever they are, will probably be important. So what is a differential manifold? The name should suggest the answer: they are spaces in which we know what differentiation is supposed to mean. I actually prefer the term *smooth manifold*, so that is what I will use going forward, though this will often just get shortened to *manifold*.

In practice, the idea is to leverage the fact that we already (hopefully!) know how to do calculus in \mathbb{R}^n (this is exactly what MATH 261 is all about), and to translate those techniques to more general spaces. The key insight here is that differentiation is a local operation: to compute a derivative at a point (whether it’s a gradient, curl, divergence, directional derivative, whatever), you really only need to know what’s going on in a tiny open neighborhood of that point.

So to get a space on which we can compute derivatives, it’s enough to have a space which is “locally Euclidean” or “locally like \mathbb{R}^n ,” and this is what manifolds are. Roughly speaking, this means that around any point in a manifold you can find a small open set which looks just like an open set in some \mathbb{R}^n , and then we can do calculus on the manifold by translating in a neighborhood of a point to the corresponding set in \mathbb{R}^n , where we know what to do.

This is all to say that the point of defining manifolds in the way we are about to (which is extremely non-obvious and unintuitive!) is that these are precisely the spaces in which a suitable generalization of multivariable calculus makes sense.

So what does “locally like \mathbb{R}^n ” actually mean? Here’s a standard definition:

Definition 1.1.1. A *smooth manifold of dimension n* is a Hausdorff, second-countable topological space M together with a family of injective maps $\phi_\alpha : U_\alpha \rightarrow M$ from open sets $U_\alpha \subseteq \mathbb{R}^n$ so that:

- (i) $\bigcup_\alpha \phi_\alpha(U_\alpha) = M$ (that is, the images of the maps ϕ_α cover all of M);
- (ii) For any α, β so that $\phi_\alpha(U_\alpha) \cap \phi_\beta(U_\beta) = W \neq \emptyset$, the sets $\phi_\alpha^{-1}(W)$ and $\phi_\beta^{-1}(W)$ are open sets in \mathbb{R}^n and the maps $\phi_\beta^{-1} \circ \phi_\alpha$ and $\phi_\alpha^{-1} \circ \phi_\beta$ (when restricted to these open sets) are smooth.
- (iii) The family $\{(U_\alpha, \phi_\alpha)\}$ is maximal with respect to (i) and (ii).

The pairs (U_α, ϕ_α) are called *coordinate charts* and the (maximal) collection $\{(U_\alpha, \phi_\alpha)\}$ is called an *atlas*.

See [Figure 1.1](#) for a visualization of (ii), showing the map $\phi_\beta^{-1} \circ \phi_\alpha$ from $\phi_\alpha^{-1}(W) \subseteq \mathbb{R}^n$ to $\phi_\beta^{-1}(W) \subseteq \mathbb{R}^n$.



Figure 1.1: Transition maps.

Remark 1.1.2. In (ii) above, “smooth” means “infinitely differentiable” or, in shorthand, C^∞ . What I’m calling *smooth manifolds* are sometimes also called *C^∞ manifolds*. More generally we can talk about C^α manifolds for any integer $\alpha \geq 0$, where we just modify (ii) to require the maps $\phi_\beta^{-1} \circ \phi_\alpha$ and $\phi_\alpha^{-1} \circ \phi_\beta$ to be C^α .¹ In the special case $\alpha = 0$, this is just a requirement that these maps be continuous, and C^0 manifolds are often called *topological manifolds*.

Example 1.1.3. The maximal family containing $(\mathbb{R}^n, \text{id})$ makes \mathbb{R}^n into a smooth manifold.

Of course, most interesting manifolds are not \mathbb{R}^n , but the idea of (i) from [Definition 1.1.1](#) is that you can cover any manifold by a bunch of little open sets (namely, the $\phi_\alpha(U_\alpha)$) that are essentially identical to open sets in \mathbb{R}^n (namely, the U_α),² so you can essentially do any local calculation in \mathbb{R}^n . It’s also very important that the n is always the same here: if $m \neq n$, we’re not allowed to have some points with neighborhoods that look like \mathbb{R}^m and some other points whose neighborhoods look like \mathbb{R}^n .

If the idea is to use the coordinate charts to transport calculations from the manifold to \mathbb{R}^n , then a thing you should be very worried about is that, if a point lies in two different charts, then there are two different ways to do this and they might not be compatible. This is the point of (ii): whether you do your calculations in U_α or U_β , the two are related by a smooth map, so you can easily translate between the two calculations using the change-of-variables formula. Indeed, the usual English-language gloss of (ii) is that “transition functions are smooth.”

¹Recall that a continuous map is C^α if it has α continuous derivatives.

²In topological terms, U_α and $\phi_\alpha(U_\alpha)$ are homeomorphic.

Finally, (iii) is a technical condition that in practice is not important. The point of it is simply to ensure uniqueness: if you had a collection of coordinate charts satisfying (i) and (ii), and I took your collection and added some new charts while still satisfying (ii), it would be kind of silly to say that you and I were talking about different manifolds. Taking maximal families gives uniqueness since your collection of charts and my collection of charts live in the same maximal family.

However, it is certainly possible to have distinct maximal collections on the same space (if there is one that is in some sense standard, then any others are sometimes called “exotic smooth structures”). At least two Fields Medals have been awarded primarily for finding examples of exotic smooth structures: to John Milnor in 1962 (for finding exotic 7-spheres [7]; it turns out there are exactly 28 distinct smooth structures on S^7 [6]) and to Simon Donaldson in 1986 (for finding exotic \mathbb{R}^4 s [1, 2, 3]; it turns out there are uncountably many distinct smooth structures on \mathbb{R}^4 [8]). It remains an open problem called the *smooth 4-dimensional Poincaré conjecture* whether there are non-standard differentiable structures on S^4 .

Remark 1.1.4. We often mimic the \mathbb{R}^n notation and indicate the dimension of a manifold M with a superscript; i.e., M^n means that M is an n -dimensional manifold, not that we are taking the Cartesian product $M \times M \times \dots \times M$.

Example 1.1.5. S^n the unit sphere in \mathbb{R}^{n+1} is a manifold. Specifically, I claim that the maximal family containing $\{(R^n, \phi_N), (R^n, \phi_S)\}$ makes S^n into an n -dimensional manifold, where ϕ_N and ϕ_S are inverse stereographic projection from the north and south poles, respectively.

Specifically, with $\vec{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$, define

$$\phi_N(\vec{x}) := \frac{1}{1 + \|\vec{x}\|^2} (2x_1, \dots, 2x_n, -1 + \|\vec{x}\|^2)$$

and

$$\phi_S(\vec{x}) := \frac{1}{1 + \|\vec{x}\|^2} (2x_1, \dots, 2x_n, 1 - \|\vec{x}\|^2).$$

Then ϕ_N is the map that sends $\vec{x} \in \mathbb{R}^n$ to the point on the sphere which lies on the line segment connecting $(x_1, \dots, x_n, 0) \in \mathbb{R}^{n+1}$ to the north pole $(0, \dots, 0, 1) \in S^n \subset \mathbb{R}^{n+1}$; see Figure 1.2.

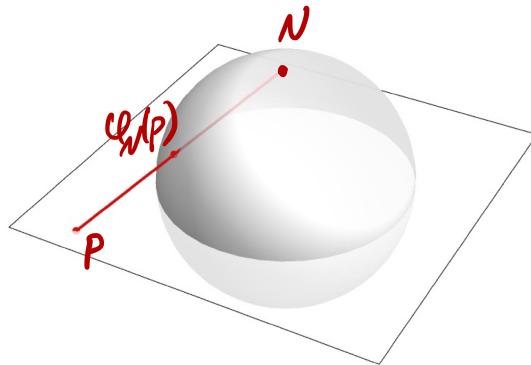


Figure 1.2: Inverse stereographic projection.

To prove the claim, we need to show that (i) and (ii) from Definition 1.1.1 are satisfied ((iii) is automatically satisfied, since we’re taking the maximal family containing $\{(\mathbb{R}^n, \phi_N), (\mathbb{R}^n, \phi_S)\}$).

If we let $N = (0, \dots, 0, 1)$ be the north pole and $S = (0, \dots, 0, -1)$ the south pole, then $\phi_N(\mathbb{R}^n) = S^n \setminus \{N\}$ and $\phi_S(\mathbb{R}^n) = S^n \setminus \{S\}$ and the union is all of S^n , so (i) is satisfied.

For (ii), observe that

$$W = \phi_N(\mathbb{R}^n) \cap \phi_S(\mathbb{R}^n) = S^n \setminus \{N, S\},$$

so

$$\phi_N^{-1}(W) = \mathbb{R}^n \setminus \{0\},$$

which is certainly open, and likewise for $\phi_S^{-1}(W) = \mathbb{R}^n \setminus \{0\}$. So we need to verify that $\phi_N^{-1} \circ \phi_S$ and $\phi_S^{-1} \circ \phi_N$ are smooth as functions on $\mathbb{R}^n \setminus \{0\}$.

The inverse of ϕ_N is stereographic projection

$$\phi_N^{-1}(\vec{y}) := \frac{1}{1 - y_{n+1}}(y_1, \dots, y_n)$$

(check this!) so we see that

$$(\phi_N^{-1} \circ \phi_S)(\vec{x}) = \frac{1}{\|\vec{x}\|^2} \vec{x}$$

is reflection through the unit sphere in \mathbb{R}^n , which is smooth away from the origin. And similarly for $\phi_S^{-1} \circ \phi_N$.

As already mentioned, the point of manifolds is that they are spaces in which we can do calculus, so we should be able to say what it means for a map between manifolds to be differentiable. Hopefully it's already starting to become clear what the strategy is: we can talk about differentiability at a point, and then both the point in the domain and the point it maps to in the range lie in coordinate charts that are like open sets in Euclidean spaces. So then locally our map just looks like a map between Euclidean spaces, where we already know what it means for a map to be differentiable.

Definition 1.1.6. Let M^m and N^n be manifolds. A continuous map $f: M \rightarrow N$ is *differentiable* at $p \in M$ if, given a coordinate chart $\psi: V \subseteq \mathbb{R}^n \rightarrow N$ containing $f(p)$, there exists a coordinate chart $\phi: U \subseteq \mathbb{R}^m \rightarrow M$ containing p so that $f(\phi(U)) \subseteq \psi(V)$ and

$$\psi^{-1} \circ f \circ \phi: U \subseteq \mathbb{R}^m \rightarrow \mathbb{R}^n$$

is differentiable at $\phi^{-1}(p)$ (see Figure 1.3). The map f is differentiable on an open set in M if it is differentiable at every point in that set.

Example 1.1.7. Consider the antipodal map $\alpha: S^n \rightarrow S^n$ given by $\alpha(\vec{y}) = -\vec{y}$. Then, so long as \vec{y} is not the south pole, $\alpha(\vec{y}) = -\vec{y} \in \phi_N(\mathbb{R}^n)$ so we can take $\psi = \phi_N$ and $V = \mathbb{R}^n$. Moreover, $\vec{y} \in \phi_S(\mathbb{R}^n)$ so we can $\phi = \phi_S$ and $U = \mathbb{R}^n$, since $\alpha(\phi_S(\mathbb{R}^n)) = S^n \setminus \{N\} = \phi_N(\mathbb{R}^n)$. Then a straightforward calculation shows that

$$(\phi_N^{-1} \circ \alpha \circ \phi_S)(\vec{x}) = -\vec{x},$$

which is definitely differentiable everywhere (as a map $\mathbb{R}^n \rightarrow \mathbb{R}^n$).

Of course, if $\vec{y} = S$, we can swap the roles of ϕ_N and ϕ_S in the above, and we conclude that α is differentiable everywhere on S^n .

While we've given the definition of a manifold and of a differentiable map in this section, we generally try to use them directly as little as possible. They are hard to handle and fairly unintuitive, so we will quickly be looking for alternative ways of characterizing manifolds and differentiable maps.

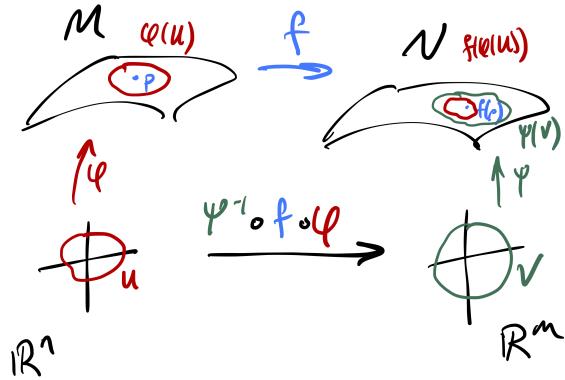


Figure 1.3: Locally converting a map between manifolds to a map between open sets in Euclidean spaces, where we know what differentiability means.

1.2 Tangent Vectors

Notice that there's nothing in [Definition 1.1.1](#) that says that a manifold has to live inside some bigger Euclidean space. This is in contrast to a typical undergraduate differential geometry course (like [MATH 474](#) here at CSU), which is typically focused on surfaces in \mathbb{R}^3 .

Of course, many manifolds (like spheres) do naturally live in some Euclidean space, and it turns out that the Whitney embedding theorem [9, 10] guarantees that all manifolds *can* be embedded in some Euclidean space, but this embedding is not necessarily going to be pleasant to work with. If you've encountered them before, it is often much easier to work with projective spaces and Grassmannians *without* embedding them anywhere in particular.

Since our manifolds don't necessarily live inside a Euclidean space, we have to be a bit careful about what a tangent vector at a point is supposed to be. In particular, a tangent vector at a point lives in a different universe than the point itself: the point is a point in the manifold, but the tangent vector does *not* live on the manifold. This is clear even on a sphere in Euclidean space: the tangent vector to a point on the sphere does not live on the sphere! However, in Euclidean space we can kind of cheat and think of both the point and the tangent vector as both living inside the ambient Euclidean space. In general, we can't get away with this.

Now even in Euclidean space you have to be a little careful with this mixing of point and tangent vector: the tangent vector can't be any *arbitrary* vector in the ambient Euclidean space: it has to lie in the *tangent space* at the point, which we usually visualize as some plane which is tangent to the sphere at the point. But if we're not in Euclidean space, things are even worse: if there's supposed to be some subspace which is "tangent at a point," where does it even live? Surely not in the manifold itself, but we're not thinking of the manifold as sitting inside some bigger space, so there's no "outside" where it can be.

This is a surprisingly nontrivial issue, requiring us to construct some abstract vector space which doesn't really live anywhere in particular. The construction is fairly non-obvious, and feels like a sneaky trick the first few times you encounter it. This is one of those situations where it seems like you're turning a concept inside-out; at least for me, the first time I encountered the following way of thinking, it made me feel uncomfortable in a similar way to when I first encountered the natural embedding of a vector space into its double dual. I will say that at some point my brain switched from "this is weird and awkward" to "this is

obviously the right way to do it" and I think most differential geometers have had similar experiences, so this is something you can eventually develop intuition for.

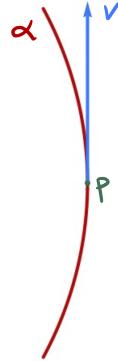
Given that preamble, what's the idea? We're going to work by analogy with a way of thinking about vectors in \mathbb{R}^n that's slightly different from what you may be used to. In words, we'll identify a vector $\vec{v} \in \mathbb{R}^n$ with the operator on differentiable functions which gives the directional derivative (in the direction of \vec{v}) of a function.

That's a little vague, so let's try to characterize a tangent vector \vec{v} to a point $p \in \mathbb{R}^n$ in this way (here I'm using the notation p rather than \vec{p} , because I'm just thinking of p as a point in a manifold, not as an element of a vector space). Given \vec{v} , I claim we can find some smooth curve $\alpha : (-\epsilon, \epsilon) \rightarrow \mathbb{R}^n$ with $\alpha(0) = p$ and $\alpha'(0) = \vec{v}$; see [Figure 1.4](#). In coordinates, if

$$\alpha(t) = (x_1(t), \dots, x_n(t)) \quad \text{for } t \in (-\epsilon, \epsilon),$$

where the coordinate functions $x_i : (-\epsilon, \epsilon) \rightarrow \mathbb{R}$ are themselves smooth, then

$$\alpha'(0) = (x'_1(0), \dots, x'_n(0)) = \vec{v}.$$



[Figure 1.4: A curve through \$p\$ with velocity \$\vec{v}\$ at \$p\$.](#)

(In this specific case, we could take $\alpha(t) = p + t\vec{v}$, but it turns out not to matter which curve satisfying $\alpha(0) = p$ and $\alpha'(0) = \vec{v}$ we take.)

Now, say $f : U \rightarrow \mathbb{R}$ is differentiable, where $U \subseteq \mathbb{R}^n$ is some neighborhood of p . Then the directional derivative of f at p in the direction of \vec{v} is

$$\frac{d(f \circ \alpha)}{dt} \Big|_{t=0} = \sum_{i=1}^n \frac{\partial f}{\partial x_i} \Big|_p \frac{dx_i}{dt} \Big|_{t=0} = \left(\sum_{i=1}^n x'_i(0) \frac{\partial}{\partial x_i} \right) f$$

by the Chain Rule.

On the right hand side above, we've written the directional derivative as an operator $L = \sum_{i=1}^n x'_i(0) \frac{\partial}{\partial x_i}$ acting on f . Moreover, this operator depends uniquely on \vec{v} and is a *linear derivation*, meaning that:

- (i) $L(f + \lambda g) = L(f) + \lambda L(g)$ for all f and g differentiable in a neighborhood of p and all $\lambda \in \mathbb{R}$;

(ii) $L(fg) = f(p)L(g) + g(p)L(f)$ (i.e., the Product Rule).

If it's not clear to you, it's worth looking at the above and convincing yourself that the operator L didn't really depend on the choice of α : it really only depends on p and \vec{v} (you may need to go back to the justification from MATH 517 that the directional derivative is well-defined).

The upshot is that, given a point p and a tangent vector \vec{v} at p , we get a directional derivative operator L . And, conversely, if we know how to compute a directional derivative, we (at least implicitly) know the direction, so this is really a bijective correspondence.

The benefit of thinking in this way is that we can talk about differential operators like L on any manifold; after all, manifolds locally look like \mathbb{R}^n by definition. So, with that long preamble in mind, here's the definition of a tangent vector:

Definition 1.2.1. Let M^n be a manifold. A smooth function $\alpha: (-\epsilon, \epsilon) \rightarrow M$ is a (smooth) curve in M . Suppose $\alpha(0) = p \in M$ and let \mathcal{D}_p be the set of functions on M that are differentiable in a neighborhood of p . The *tangent vector* to a curve α at $t = 0$ is a function $\alpha'(0): \mathcal{D}_p \rightarrow \mathbb{R}$ given by

$$\alpha'(0)f := \frac{d(f \circ \alpha)}{dt} \Big|_{t=0} = (f \circ \alpha)'(0).$$

A *tangent vector at p* is the tangent vector at $t = 0$ of some curve $\alpha: (-\epsilon, \epsilon) \rightarrow M$ with $\alpha(0) = p$.

The set of all tangent vectors at p is the *tangent space* to M at p , denoted $T_p M$.

This is a nice coordinate-free way of defining things, but it's not very useful for computations. We usually *do* want to work in coordinates for computations (and certainly for any computations that we want to do on the computer), so let's see what all this means in local coordinates.

Say that (U, ϕ) is a coordinate chart containing $p \in M$ so that $\phi(\vec{0}) = p$, that $\alpha: (-\epsilon, \epsilon) \rightarrow M$ is smooth with $\alpha(0) = p$, and that f is a differentiable function in a neighborhood of p . Then

$$(\phi^{-1} \circ \alpha)(t) = (x_1(t), \dots, x_n(t))$$

for some smooth functions $x_1, \dots, x_n: (-\epsilon, \epsilon) \rightarrow \mathbb{R}$. Then

$$\alpha'(0)f = \frac{d(f \circ \alpha)}{dt} \Big|_{t=0} = \frac{d}{dt}(f \circ \phi(x_1(t), \dots, x_n(t))) \Big|_{t=0} = \sum_{i=1}^n x'_i(0) \frac{\partial f}{\partial x_i} \Big|_{\vec{0}} = \left(\sum_{i=1}^n x'_i(0) \frac{\partial}{\partial x_i} \Big|_{\vec{0}} \right) f,$$

where I'm using a very common abuse of notation in the second expression to think of f as being a function on the coordinate chart U , and hence a function of coordinates x_1, \dots, x_n (of course, it's really $f \circ \phi$ which is a function of x_1, \dots, x_n , as we see in the middle expression).

This all means that we can write the tangent vector $\alpha'(0) \in T_p M$ in local coordinates as

$$\alpha'(0) = \sum_{i=1}^n x'_i(0) \left(\frac{\partial}{\partial x_i} \Big|_{\vec{0}} \right).$$

Since the $x'_i(0)$ are just scalars, what we're doing here is writing $\alpha'(0)$ in terms of the *local coordinate basis* $\left\{ \left(\frac{\partial}{\partial x_1} \Big|_{\vec{0}}, \dots, \frac{\partial}{\partial x_n} \Big|_{\vec{0}} \right) \right\}$ for $T_p M$ associated to the chart (U, ϕ) .

Remark 1.2.2. In practice, we will almost always drop the subscript $\vec{0}$ and just write the basis as $\left\{\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_n}\right\}$ and generic tangent vectors as $\sum_{i=1}^n a_i \frac{\partial}{\partial x_i}$.

Exercise 1.2.3. Suppose a point $p \in M$ lies in two different coordinate charts. How are the two different local coordinate bases related?

It's extremely important to keep in mind that, if p and q are distinct points on M , then the tangent spaces $T_p M$ and $T_q M$ are *completely different vector spaces* that in principle have nothing to do with each other (they're both vector spaces of the same dimension, and hence abstractly isomorphic, but that's essentially all we can know without more detailed information about the geometry of M).

Nonetheless, we often want to talk about all tangent spaces at once, so we smush them together:

Definition 1.2.4. The *tangent bundle* of a manifold M , denoted TM , is the (disjoint) union of tangent spaces

$$TM := \bigsqcup_{p \in M} T_p M.$$

Likewise, if $(T_p M)^*$ is the dual of $T_p M$, then the *cotangent bundle* is the union of cotangent spaces

$$T^*M := \bigsqcup_{p \in M} (T_p M)^*.$$

Notice that there are natural projections $\pi : TM \rightarrow M$ and $\tilde{\pi} : T^*M \rightarrow M$ which just record the base point; in other words, π sends a tangent vector at a point to the point (formally, TM and T^*M are *vector bundles*, and this projection is part of their definition).

Theorem 1.2.5. If M is an n -dimensional manifold, then TM and T^*M are $2n$ -dimensional manifolds.

Exercise 1.2.6. Prove Theorem 1.2.5.

T^*M is, in some sense, the most basic example of a *symplectic manifold*. In physics and dynamical systems language, if M is the *configuration space*³ of a (classical) system, then T^*M is *phase space* or *position-momentum space*: this is the natural setting of Hamiltonian mechanics.

Example 1.2.7. $T\mathbb{R}^n \cong \mathbb{R}^n \times \mathbb{R}^n \cong \mathbb{R}^{2n}$.

Example 1.2.8. $TS^1 \cong S^1 \times \mathbb{R}$, the infinite cylinder.

Example 1.2.9. TS^2 is a *non-trivial* bundle over S^2 , meaning that it is *not* homeomorphic to $S^2 \times \mathbb{R}^2$. In particular, one can show that US^2 , which is the *unit tangent bundle* (the subset of the tangent bundle consisting only of unit tangent vectors) is homeomorphic to $\text{SO}(3) \cong \mathbb{RP}^3$, the real projective space, which is a circle bundle over S^2 different from the trivial circle bundle $S^2 \times S^1$. (For those that have taken graduate topology, $\pi_1(S^2 \times S^1) \cong \mathbb{Z}$, whereas $\pi_1(\mathbb{RP}^3) \cong \mathbb{Z}/2\mathbb{Z}$.)

³Meaning it records the positions of particles; for example if we're doing dynamics of n points on the circle, the configuration space is the n -torus $S^1 \times \dots \times S^1 = (S^1)^n$.

1.3 The Differential

In multivariable calculus, any differentiable map $f: \mathbb{R}^m \rightarrow \mathbb{R}^n$ has a corresponding differential df .⁴ While in a multivariable calculus setting we often think of df as a map $\mathbb{R}^m \rightarrow \mathbb{R}^n$ as well (meaning it would have the same domain and codomain as f), or even more concretely as an $n \times m$ matrix, this is misleading.

In fact, if you were reading the first sentence of the previous paragraph very carefully, you may have noticed that it was not quite correct. Really, I should have said that any differentiable map has a corresponding differential *at a point p* . There is no single “differential” for a non-linear map: we get different differentials in the neighborhood of each point in the domain.

So more accurately, if p is in the domain of f , then we have a differential df_p at p . Now, what is the domain of df_p ? Recall, df_p is supposed to be the best linear approximation of f at p , so we’re supposed to think of the inputs to df_p as being slight perturbations of p . In other words, the domain of df_p is really the tangent space $T_p \mathbb{R}^m$ (which is in a very precise sense the space of *infinitesimal* perturbations of p).

Now, $T_p \mathbb{R}^m$ is isomorphic to \mathbb{R}^m , but not just in some abstract sense: the tangent bundle $T \mathbb{R}^m$ is trivial, meaning that $T \mathbb{R}^m \cong \mathbb{R}^m \times \mathbb{R}^m$, and a concrete trivialization is given by parallel translating a tangent vector from being based at p to being based at the origin. This is the sense in which we can think of the domain of each df_p as being the “same” \mathbb{R}^m .

But if M is some more general m -dimensional manifold, then, while any tangent space $T_p M$ is still abstractly isomorphic to \mathbb{R}^m , the tangent bundle is likely not to be trivial, so we can’t make any general identifications of different tangent spaces with the same \mathbb{R}^m . So if we want to generalize the notion of differential to manifolds, we need to be careful, even when thinking about functions on \mathbb{R}^m , to maintain the distinction between \mathbb{R}^m (thought of an m -dimensional manifold) and $T_p \mathbb{R}^m$.

And what about the codomain of df_p ? Again, we should think more conceptually about what the differential really does. Like all derivatives, the differential tells us something about how the change in an input to a function changes the output of the function. So the inputs to the differential will be tangent vectors at some point in the domain of f (recording an infinitesimal change to the input, or equivalently an initial position [the point] and velocity [the tangent vector] of some path), and the outputs will be tangent vectors at the corresponding point in the range (recording the infinitesimal change in the output). Notice that $T_q \mathbb{R}^n$ is again isomorphic with \mathbb{R}^n by parallel translation.

In other words, given a differentiable map $f: \mathbb{R}^m \rightarrow \mathbb{R}^n$ and a point p in the domain of f , the differential of f at p is a map $df_p: T_p \mathbb{R}^m \rightarrow T_{f(p)} \mathbb{R}^n$, where domain and range are usually identified with \mathbb{R}^m and \mathbb{R}^n in a standard way. This, then, is the way of thinking about differentials which generalizes.

Definition 1.3.1. Suppose M^m and N^n are smooth manifolds and $f: M \rightarrow N$ is smooth. For each $p \in M$, define the *differential* of f at p , denoted $df_p: T_p M \rightarrow T_{f(p)} N$, as follows: For any $v \in T_p M$, choose a smooth curve $\alpha: (-\epsilon, \epsilon) \rightarrow M$ such that $\alpha(0) = p$ and $\alpha'(0) = v$. Let $\beta = f \circ \alpha$, which is a smooth curve in N . Then

$$df_p(v) := \beta'(0).$$

Lemma 1.3.2. *This is a well-defined linear map.*

Exercise 1.3.3. Prove Lemma 1.3.2.

In the special case that $f: M \rightarrow \mathbb{R}$ is a smooth function, the differential of f applied to a tangent vector v is the same thing as computing the derivative of f in the direction of v :

⁴You may have encountered this under the name *Jacobian* rather than *differential*; in the case $n = 1$, this is [more or less] the gradient of the function.

Lemma 1.3.4. Suppose M is a manifold, $p \in M$, $v \in T_p M$, and that $f: M \rightarrow \mathbb{R}$ is differentiable in a neighborhood of p . Then

$$vf = df_p v.$$

Proof. This is just a matter of unwinding definitions. Let $\alpha: (-\epsilon, \epsilon) \rightarrow M$ be a smooth curve so that $\alpha(0) = p$ and $\alpha'(0) = v$. Then by definition the left hand side is

$$(vf)(p) = (\alpha'(0)f)(p) = (f \circ \alpha)'(0).$$

But this is exactly the definition of $df_p v$. \square

Let's see what this means in local coordinates, which is how we'll actually compute in practice. Suppose $p \in M$ and (V, ψ) is a local coordinate chart on N containing $f(p)$. By Definition 1.1.6, there exists some compatible chart (U, ϕ) on M containing p so that $f(\phi(U)) \subseteq \psi(V)$. Then the curve α on M (or, really, its restriction to $\phi(U)$) gives a corresponding curve

$$(x_1(t), \dots, x_m(t)) = \phi^{-1} \circ \alpha(t)$$

in $U \subseteq \mathbb{R}^m$, and similarly the curve β on N has a corresponding curve

$$(y_1(t), \dots, y_n(t)) = \psi^{-1} \circ \beta(t).$$

See Figure 1.5.

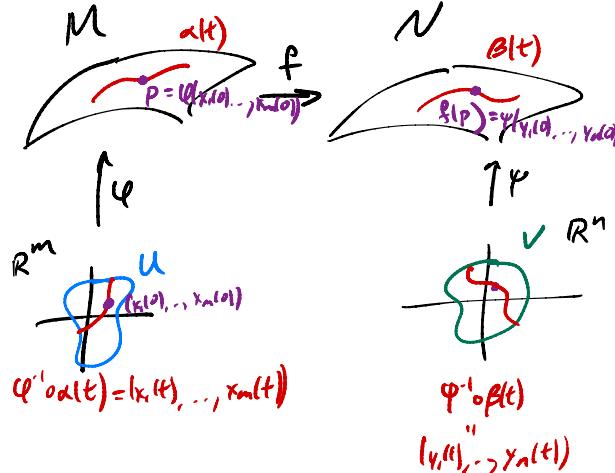


Figure 1.5: The differential in local coordinates

Of course,

$$\psi^{-1} \circ \beta(t) = \psi^{-1} \circ f \circ \alpha(t) = \psi^{-1} \circ f \circ \phi(x_1(t), \dots, x_m(t)),$$

so we can think of the y_i as functions of the x_j , and the curve in V can be written as

$$(y_1(x_1(t), \dots, x_m(t)), \dots, y_n(x_1(t), \dots, x_m(t))).$$

Working in local coordinates means doing the computation at the level of the map $\psi^{-1} \circ f \circ \phi: U \subseteq \mathbb{R}^m \rightarrow V \subseteq \mathbb{R}^n$, which has a corresponding differential

$$d(\psi^{-1} \circ f \circ \phi)_{(x_1(0), \dots, x_m(0))}: T_{(x_1(0), \dots, x_m(0))}\mathbb{R}^m \rightarrow T_{(y_1(0), \dots, y_m(0))}\mathbb{R}^n.$$

By definition, evaluating this differential on the tangent vector $(x'_1(0), \dots, x'_m(0)) \in T_{(x_1(0), \dots, x_m(0))}\mathbb{R}^m$ gives

$$\begin{aligned} \frac{d}{dt} \Big|_{t=0} (y_1(x_1(t), \dots, x_m(t)), \dots, y_n(x_1(t), \dots, x_m(t))) &= \left(\sum_{j=1}^m \frac{\partial y_1}{\partial x_j} \Big|_x \frac{dx_j}{dt} \Big|_{t=0}, \dots, \sum_{j=1}^m \frac{\partial y_n}{\partial x_j} \Big|_x \frac{dx_j}{dt} \Big|_{t=0} \right) \\ &= \left(\sum_{j=1}^m \frac{\partial y_1}{\partial x_j} \Big|_x x'_j(0), \dots, \sum_{j=1}^m \frac{\partial y_n}{\partial x_j} \Big|_x x'_j(0) \right) \\ &= \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_n}{\partial x_1} & \cdots & \frac{\partial y_n}{\partial x_m} \end{bmatrix} \begin{bmatrix} x'_1(0) \\ \vdots \\ x'_m(0) \end{bmatrix} \\ &= \begin{bmatrix} \frac{\partial y_i}{\partial x_j} \end{bmatrix}_{i,j} x'(0), \end{aligned}$$

where $\begin{bmatrix} \frac{\partial y_i}{\partial x_j} \end{bmatrix}_{i,j}$ is the $n \times m$ matrix of partials (i.e., the Jacobian).

So when we work in local coordinates, this is just standard multivariable calculus, as you would hope.

Just to reiterate, the differential (at a point) of a map between manifolds is going to input a tangent vector at a point in the domain and output a tangent vector at a point in the range. As a special case, suppose we have a smooth function on a manifold M ; that is, a smooth map $f: M \rightarrow \mathbb{R}$. Then, at a point $p \in M$, the differential $df_p: T_p M \rightarrow T_{f(p)}\mathbb{R}$. Now, the tangent bundle to \mathbb{R} is trivial, so we can identify $T_{f(p)}\mathbb{R}$ with \mathbb{R} in a canonical way, and therefore think about df_p as a linear map $T_p M \rightarrow \mathbb{R}$.

In other words, by way of this identification of $T_{f(p)}\mathbb{R}$ with \mathbb{R} , we can think of df_p as a linear functional on the vector space $T_p M$. Or, in fancier language, df_p is an element of the dual space $(T_p M)^*$. We'll come back to this later when we start talking about differential forms, but this gives a hint of why we might care about cotangent spaces and the cotangent bundle.

Just as in multivariable calculus, when we have a smooth map $f: M \rightarrow N$, we can talk about critical points and critical values, and doing so is often quite important.

Definition 1.3.5. Let $f: M \rightarrow N$ be smooth. A point $p \in M$ is a *critical point* of f if $df_p: T_p M \rightarrow T_{f(p)}N$ is not surjective; then $f(p)$ is a *critical value* of f . A point $q \in N$ which is not a critical value is a *regular value* of f .

Example 1.3.6 (Cliché Example). Consider the function f on the torus depicted in Figure 1.6. In words, f is the height function on this upright torus that just records the z -coordinate. I've shown the critical values in red, as well as a couple of regular values in green. Also, back on the torus I've shown the inverse images of the critical and regular values, as well as some tangent vectors to regular points and their images under the differential (in blue and purple), to hopefully convince you the differential really is surjective at these points.

Notice that there are basically three different types of critical points: a minimum, a maximum, and two saddle points. The differential can't possibly be surjective at a (local) minimum: there's no direction you



Figure 1.6: The height function on a torus.

could go which will cause the function to decrease, so the image of the differential has to miss an entire half of its codomain (and therefore, by linearity, everything except the origin). It's somewhat less geometrically clear that the saddle points are critical points, though in this case it's fairly easy to see once you realize the tangent plane to one of the saddle points is horizontal.

A couple of other things to notice about this picture:

- (i) The inverse images of the two critical points coming from the saddle points contain plenty of regular points: indeed, the differential is surjective at any of the points in the inverse image *except* the saddle point.
- (ii) The level sets of regular values are smooth (collections of) curves, either a single circle or two disjoint circles; in other words, they are smooth (possibly disconnected) 1-dimensional manifolds. The level sets of the critical points are either a single point (the minimum and maximum) or a wedge of two circles (like an ∞ symbol). In both cases, the result is not a smooth 1-manifold: a point is a 0-manifold and the wedge of two circles is not locally like a copy of \mathbb{R} near the point where the two circles meet (which of course is exactly the saddle point).

Remark 1.3.7. I call this a cliché example because it is the first picture everybody draws when they talk about Morse theory, which says that the topology of a manifold is in some sense encoded in the critical points of any sufficiently generic function on the manifold. Such functions are called *Morse functions*, and the function in this example is a Morse function. In particular, Morse theory tells you that if you are traversing the manifold “up” according to a Morse function, the topology only changes when you pass a critical point. In this example, the topology changes from the empty set to a disk when we pass the minimum, then from

a disk to a

(which is homotopic to a circle) when we pass the first saddle point, then to a punctured

torus (which is homotopic to a wedge of circles) when we pass the second saddle point, and then finally the puncture is filled in when we pass the maximum. Something analogous happens in general.

Example 1.3.8. Consider the function $f(x, y, z) = z^2$ on the sphere (see Figure 1.7). Again, I've shown critical values in red and orange and a regular value in green, and the corresponding level sets on the domain, as well as some tangent vectors to regular points and their images under the differential.

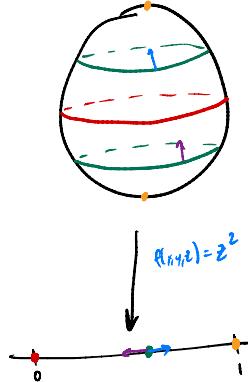


Figure 1.7: A smooth function on the sphere which does not have isolated critical points.

Some new features:

- (i) The critical points aren't all isolated: the entire equator consists of critical points (which are all global minima).
- (ii) The inverse images of critical values aren't necessarily connected: the north and south poles both map to 1.

Notice again that the level sets of regular values are (collections of) smooth curves. In this case one critical level set is also a smooth curve, and the other is not a curve at all.

(This is not a Morse function because not all critical points are isolated, but it is a *Morse–Bott function*, which is almost as good.)

After looking at these examples, hopefully the following theorem suggests itself:

Theorem 1.3.9 (Level Set Theorem). *If $m \geq n$, $f: M^m \rightarrow N^n$ is smooth, and $q \in N$ is a regular value of f , then $f^{-1}(q)$ is a smooth submanifold of M of dimension $m - n$.*

This theorem is basically an application of the Inverse Function Theorem, but before we work on proving it, let's look at some more examples.

Example 1.3.10. Define $f: \mathbb{R}^n \rightarrow \mathbb{R}$ by $f(x_1, \dots, x_n) = x_1^2 + \dots + x_n^2$. Then

$$df_{(x_1, \dots, x_n)} = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \cdots & \frac{\partial f}{\partial x_n} \end{bmatrix} = \begin{bmatrix} 2x_1 & \cdots & 2x_n \end{bmatrix}$$

only fails to be full rank at the origin, so 0 is the only critical value, and for all $r > 0$ the level set $f^{-1}(r)$ is a smooth manifold of dimension $n - 1$. But $f^{-1}(r)$ is nothing but the sphere of radius \sqrt{r} , so we've just given an alternative (and somehow conceptually prior) proof that spheres are manifolds.

Example 1.3.11 (Relevant to frame theory). Let $\text{Mat}_{d \times N}(\mathbb{C})$ be the space of $d \times N$ complex matrices. This is trivially a $2dN$ -dimensional manifold, since $\text{Mat}_{d \times N}(\mathbb{C}) \cong \mathbb{C}^{dN}$, which is just a copy of \mathbb{R}^{2dN} if you ignore the complex structure. Now, let $\mathcal{H}(d)$ be the space of $d \times d$ Hermitian matrices (i.e., $d \times d$ complex matrices A so that $A^* = A$, where A^* is the conjugate transpose of A) and define the map $\Phi: \text{Mat}_{d \times N}(\mathbb{C}) \rightarrow \mathcal{H}(d)$ by

$$\Phi(A) = AA^*.$$

(In frame theory, we would think of $A \in \text{Mat}_{d \times N}(\mathbb{C})$ as a frame and AA^* as the corresponding frame operator.)

I claim that the identity matrix $I_{d \times d} \in \mathcal{H}(d)$ is a regular value of Φ (**Exercise:** Prove this); assuming this, the theorem tells us that $\Phi^{-1}(I_{d \times d})$ is a smooth manifold of dimension

$$2dN - d^2 = d(2N - d).$$

(Notice that $A \in \Phi^{-1}(I_{d \times d})$ means that the rows of A are Hermitian orthonormal. Since each row is a vector in \mathbb{C}^N , this means that we can think of $\Phi^{-1}(I_{d \times d})$ as the space of all [ordered] d -tuples of Hermitian orthonormal vectors in \mathbb{C}^N . This space is an example of a *Stiefel manifold*, and usually denoted $\text{St}_d(\mathbb{C}^N)$ or $V_d(\mathbb{C}^N)$. In frame theory, $\Phi^{-1}(I_{d \times d})$ is precisely the space of *Parseval frames*.)

Example 1.3.12. Let $N = d$ in the previous example. Then $\Phi^{-1}(I_{d \times d})$ is a smooth manifold of dimension d^2 that consists of those $d \times d$ complex matrices A so that $AA^* = I_{d \times d}$. But this is precisely the unitary group $U(d)$! So we've proved that $U(d)$ is a d^2 -dimensional manifold for any d .

Remark 1.3.13. You can play the same game over \mathbb{R} to show that real Stiefel manifolds and the orthogonal group $O(d)$ are manifolds.

1.4 Immersions and Embeddings

Let's work up to proving [Theorem 1.3.9](#), including defining some more terminology.

Definition 1.4.1. Suppose $f: M \rightarrow N$ is smooth. Then f is an *immersion* if df_p is injective for all $p \in M$ (note that this implies $\dim(M) \leq \dim(N)$).

If f is also a homeomorphism onto its image (continuous bijection with continuous inverse), then f is an *embedding*. The image of an embedding is a *submanifold*.

If f is bijective and f^{-1} is smooth, then f is a *diffeomorphism*. f is called a *local diffeomorphism* at $p \in M$ if there exist neighborhoods U of p and V of $f(p)$ so that $f: U \rightarrow V$ is a diffeomorphism.

Finally, we say that f is a *submersion* if df_p is surjective for all $p \in M$.

Example 1.4.2. Define $\alpha: \mathbb{R} \rightarrow \mathbb{R}^2$ by $\alpha(t) = (t^3, t^2)$ (see [Figure 1.8](#)).

This is not an immersion because $d\alpha_0 = \begin{bmatrix} \alpha'_1(0) \\ \alpha'_2(0) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ has rank 0. Intuitively, the problem is that the velocity is zero when $t = 0$, even though the map overall is continuous and injective.

Example 1.4.3. Define $\beta: \mathbb{R} \rightarrow \mathbb{R}^2$ by $\beta(t) = (t^3 - 4t, t^2 - 4)$, which is a deformation of the previous example (see [Figure 1.9](#)).

Now the differential is given by

$$d\beta_t = \begin{bmatrix} 3t^2 - 4 \\ 2t \end{bmatrix},$$



Figure 1.8: The image of the map α from Example 1.4.2. The fact that it has a cusp suggests that it will fail to be an immersion, despite being injective and smooth.

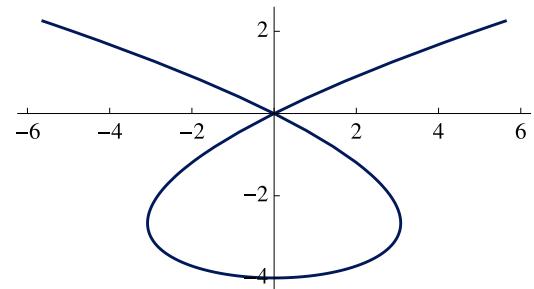


Figure 1.9: The image of the map β from Example 1.4.3. This map is an immersion, but not an embedding.

which is never the zero matrix since the second entry being 0 implies $t = 0$, and hence that the first entry is -4 . Hence, $d\beta_t$ is always rank 1, and hence always injective, so β is an immersion.

However, β is not an embedding because it is not injective: $\beta(-2) = (0, 0) = \beta(2)$, so there is a double point at the origin.

Example 1.4.4. Suppose (U, ϕ) is a local coordinate chart on a manifold M containing a point $p \in M$. Then $\phi: U \subset \mathbb{R}^n \rightarrow M$ is a local diffeomorphism at $\phi^{-1}(p) \in U$. This just follows from the definition of a coordinate chart (Definition 1.1.1): ϕ maps U bijectively onto $\phi(U)$, which is an open neighborhood of p , and both ϕ and ϕ^{-1} are smooth. (To be really pedantic, (U, id) gives a global coordinate chart for the n -manifold U , and then, following Definition 1.1.6, $\phi: U \rightarrow \phi(U)$ is smooth because $\phi^{-1} \circ \phi \circ \text{id} = \text{id}$ is smooth everywhere on U . Similarly, $\phi^{-1}: \phi(U) \rightarrow U$ is smooth because $\text{id} \circ \phi^{-1} \circ \phi = \text{id}$ is smooth everywhere on U .)

If $f: M \rightarrow N$ is a local diffeomorphism at $p \in M$, then $df_p: T_p M \rightarrow T_{f(p)} N$ is a linear isomorphism (i.e., invertible linear map). Perhaps somewhat surprisingly, the converse is also true. This is the appropriate generalization of the Inverse Function Theorem to the manifold setting, and the proof essentially involves applying the Inverse Function Theorem in local coordinates:

Proposition 1.4.5. Suppose $f: M \rightarrow N$ is smooth. If $p \in M$ and $df_p: T_p M \rightarrow T_{f(p)} N$ is an isomorphism, then f is a local diffeomorphism at p .

Proof. We are going to apply the Inverse Function Theorem (Theorem 1.4.7 below) to $\psi^{-1} \circ f \circ \phi: \phi(U) \rightarrow \psi(V)$, where (U, ϕ) and (V, ψ) are the coordinate charts guaranteed to exist by the definition of smoothness (Definition 1.1.6).

Notice that, by the Chain Rule,

$$d(\psi^{-1} \circ f \circ \phi)_{\phi^{-1}(p)} = d\psi_{f(p)}^{-1} \circ df_p \circ d\phi_{\phi^{-1}(p)}.$$

But then we already know that $d\phi$ and $d\psi^{-1}$ are isomorphisms (since coordinate charts are local diffeomorphisms [Example 1.4.4]), so df_p being an isomorphism implies that $d(\psi^{-1} \circ f \circ \phi)_{\phi^{-1}(p)}$ is also an isomorphism.

But then the Inverse Function Theorem implies $\psi^{-1} \circ f \circ \phi$ is a local diffeomorphism at $\phi^{-1}(p)$. Since ψ and ϕ^{-1} are local diffeomorphisms (again, Example 1.4.4) and compositions of local diffeomorphisms are local diffeomorphisms, it follows that

$$\psi \circ (\psi^{-1} \circ f \circ \phi) \circ \phi^{-1} = (\psi \circ \psi^{-1}) \circ f \circ (\phi \circ \phi^{-1}) = f$$

is also a local diffeomorphism. □

Here's a statement of the Inverse Function Theorem in the language of differentials and local diffeomorphisms. It is equivalent to the usual statement you would see in a multivariable calculus or analysis course.

Exercise 1.4.6. Convince yourself of the previous sentence.

Theorem 1.4.7 (Inverse Function Theorem). *If $F: U \subseteq \mathbb{R}^n \rightarrow \mathbb{R}^n$ is given by $(x_1, \dots, x_n) \mapsto (y_1, \dots, y_n)$, then $dF_p = \left[\frac{\partial y_i}{\partial x_j} \Big|_p \right]_{i,j}$ being nonsingular at $p \in U$ implies that F is a local diffeomorphism at p .*

We're now ready to prove the Level Set Theorem (Theorem 1.3.9):

Proof of Theorem 1.3.9. Recall that, in the statement, we're assuming that $f: M \rightarrow N$ is smooth and that $q \in N$ is a regular value of f , and the goal is to show that $f^{-1}(q)$ is a smooth submanifold of M of dimension $m - n$.

Suppose (U, ϕ) is a coordinate chart centered at $p \in f^{-1}(q)$ ⁵ and (V, ψ) is a chart at q . See Figure 1.10. Then define the map

$$g := \psi^{-1} \circ f \circ \phi : U \rightarrow \mathbb{R}^n.$$

By assumption,

$$dg_{\vec{0}} = d(\psi^{-1} \circ f \circ \phi)_{\vec{0}} = (d\psi^{-1})_{f(p)} \circ df_p \circ d\phi_{\vec{0}}$$

is surjective since df_p is and the other two terms are isomorphisms.



Figure 1.10: A sketch of the key neighborhoods and maps in the proof of Theorem 1.3.9.

Hence, after a linear change of coordinates $dg_{\vec{0}}$ can be written as the $n \times m$ block matrix $\begin{bmatrix} I_{n \times n} & 0_{n \times (m-n)} \end{bmatrix}$. Define

$$G(x_1, \dots, x_m) = (g(x_1, \dots, x_n), x_{n+1}, \dots, x_m),$$

which has differential $dG_{\vec{0}} = I_{m \times m}$ in these coordinates. By the Inverse Function Theorem, G is a local diffeomorphism at $\vec{0}$ and by construction $g \circ G^{-1}$ is the standard projection onto the first n coordinates.

So in these coordinates $f^{-1}(q) \cap \phi(U) = \phi(0, \dots, 0, x_{n+1}, \dots, x_m)$, so x_{n+1}, \dots, x_m give local coordinates near p for $f^{-1}(q)$. We can do the same at any other point of $f^{-1}(q)$, so this gives a system of coordinate charts on $f^{-1}(q)$, and the transition maps on overlapping charts are smooth because they are just restrictions (or projections, if you prefer) of the corresponding transition maps for the charts on M . \square

⁵Meaning that $p = \phi(\vec{0})$.

1.5 Vector Fields

Now that we have defined tangent vectors and seen how to push them around with differentials, the next natural object to try to define is a vector field. In physical problems, a vector field might be the velocity field of a fluid flow, or an electrical or magnetic field. In both applied and pure problems, we are very often interested in vector fields that arise as the gradients of functions, whether they be Morse functions on an abstract manifold or the energy of conformations in some conformation space. In symplectic geometry we are very interested in Hamiltonian vector fields associated to functions, which are a sort of symplectic gradient. Whereas the gradient is perpendicular to level sets of the function, Hamiltonian vector fields are parallel to level sets, so the function is constant on orbits, which is desirable when the function is an energy function.

So what is a vector field? Intuitively, it's just what you would expect: a choice of tangent vector at each point in the manifold. As with everything in this class, we're mostly interested in *smooth* vector fields, meaning that the tangent vector should in some sense vary smoothly as you move around the manifold.

The preceding sentence hopefully makes sense on an intuitive level, but you should stop and think about how you might try to make it into a rigorous definition. Unless you've already seen the forthcoming definition before, it's probably not so easy.

The problem is that, as alluded to previously, different tangent spaces are not directly comparable. How do you compare $v_1 \in T_{p_1} M$ with $v_2 \in T_{p_2} M$ and, ideally, put them into some difference quotient?

Since we know how to compare tangent spaces in Euclidean space (by parallel translating), one strategy is to work in local coordinates and require the tangent vectors to form a smooth vector field in each local coordinate chart. But this rather unwieldy, and in any case we want to state definitions without reference to local coordinates if at all possible. Hence the following definition, which encapsulates exactly the idea above in a very concise (though admittedly kind of abstract and hard to visualize) way:

Definition 1.5.1. A *vector field* X on a smooth manifold M is a smooth section of the tangent bundle TM .

This requires some unpacking, especially since there's at least one term in this definition which has not been defined yet.

First, recall ([Theorem 1.2.5](#)) that TM is a smooth manifold, so it makes sense to talk about smoothness of a map $X: M \rightarrow TM$. So the first part of [Definition 1.5.1](#) is that a vector field is such a smooth map.

This makes sense: such a map takes as input a point in the manifold and outputs a tangent vector. But it would be nonsensical to output a tangent vector at q if the input is p : the word “section” is how we rule this out.

More precisely, recall that we have a natural projection map $\pi: TM \rightarrow M$ which maps $v \in T_p M$ to the base point p . In general, if $E \xrightarrow{\pi} B$ is a vector bundle, then a *section* of the bundle is a map $\sigma: B \rightarrow E$ so that $\pi \circ \sigma = \text{id}_B$, the identity map on B (in algebraic terms, σ is a right inverse of the projection π).

So when we say that X is a smooth section of the tangent bundle, we mean that $X: M \rightarrow TM$ is a smooth map satisfying $\pi \circ X = \text{id}_M$. In other words, that X assigns each point in M a tangent vector at that point in a smoothly-varying way. Which is exactly what a (smooth) vector field should be!

Notation. We use the notation $\mathfrak{X}(M)$ to denote the $C^\infty(M)$ -module of vector spaces on M .

Remark 1.5.2. (i) If $\phi: U \subseteq \mathbb{R}^n \rightarrow M$ is a local coordinate chart at $p \in M$, then

$$X(p) = \sum_{i=1}^n a_i(p) \frac{\partial}{\partial x_i},$$

where each $a_i: U \rightarrow \mathbb{R}$ is smooth and $\left\{ \frac{\partial}{\partial x_i} \right\}_i$ is the local coordinate basis of $T_p M$ associated to (U, ϕ) . Notice that this is exactly the local coordinate definition of a smooth vector field informally formulated above.

(ii) Since each individual tangent vector is really a directional derivative at a point, a vector field can be interpreted as a differential operator on M : it will input a smooth function and output some new smooth function which records the directional derivative at each point in the direction specified by the tangent vector at that point. In other words, we can also interpret a vector field X on a manifold M as a map $C^\infty(M) \rightarrow C^\infty(M)$ given by $f \mapsto Xf$. In local coordinates, the function Xf is given by

$$(Xf)(p) = \sum_{i=1}^n a_i(p) \left. \frac{\partial f}{\partial x_i} \right|_p,$$

which is indeed a smooth function.

Example 1.5.3. Consider the unitary group $U(d)$ and let $I = I_{d \times d}$ be the identity matrix. I claim that a choice of tangent vector at the identity actually determines a vector field on all of $U(d)$.

To see this, recall that an element of $T_I U(d)$ is a tangent vector, which is to say, the velocity of a smooth curve $\alpha: (-\epsilon, \epsilon) \rightarrow U(d)$ with $\alpha(0) = I$. If we think of $U(d)$ as living inside $\text{Mat}_{d \times d}(\mathbb{C}) \cong \mathbb{C}^{d^2} \cong \mathbb{R}^{2d^2}$, then we can think of any tangent vector $\alpha'(0)$ as a $d \times d$ complex matrix. What restrictions does being tangent to $U(d)$ place on this matrix?

Notice that, for all t , $\alpha(t) \in U(d)$, which by definition means that $\alpha(t)\alpha(t)^* = I$. Since this is the defining equation of $U(d)$, differentiating at $t = 0$ should give precisely the condition for being in $T_I U(d)$:

$$0 = \left. \frac{d}{dt} \right|_{t=0} I = \left. \frac{d}{dt} \right|_{t=0} (\alpha(t)\alpha(t)^*) = \alpha'(0)\alpha(0)^* + \alpha(0)\alpha'(0)^* = \alpha'(0) + \alpha'(0)^*$$

since $\alpha(0) = I = \alpha(0)^*$ (note that we're also using $(\alpha(t)^*)' = \alpha'(t)^*$, which is obvious if you choose a basis and write $\alpha(t)$ as a matrix, but kind of annoying to prove in a coordinate-free way).

So we see that $\alpha'(0)^* = -\alpha'(0)$. In other words, elements of $T_I U(d)$ are precisely the skew-Hermitian matrices.

Next, how do we get from a tangent vector at I to a vector field on all of $U(d)$? Well, we know how to push vector fields around by differentials of maps, so if we had a map $U(d) \rightarrow U(d)$ sending I to a specified element $U \in U(d)$, then its differential would send $\Delta \in T_I U(d)$ to something in $T_U U(d)$. Of course, there's an obvious map sending I to U : the map which left-multiplies elements of $U(d)$ by U .

In symbols, for $U \in U(d)$, define the map $L_U: U(d) \rightarrow U(d)$ to be left-multiplication by U , namely $L_U(A) := UA$. Then certainly

$$(dL_U)_I: T_I U(d) \rightarrow T_U U(d)$$

and so, for any $\Delta \in T_I U(d)$, we get a vector field $X_\Delta \in \mathfrak{X}(U(d))$ on $U(d)$ defined by

$$X_\Delta(U) := (dL_U)_I \Delta.$$

We can make this more explicit by finding a formula for $(dL_U)_I$:

Lemma 1.5.4. If $U \in \mathrm{U}(d)$ and $\Delta \in T_I \mathrm{U}(d)$, then

$$(dL_U)_I \Delta = U\Delta.$$

Exercise 1.5.5. Prove Lemma 1.5.4. The key observation here is that matrix multiplication is linear.

Since $(dL_U)_I : T_I \mathrm{U}(d) \rightarrow T_U \mathrm{U}(d)$ is full rank, it's surjective, so $T_U \mathrm{U}(d)$ consists of matrices of the form $U\Delta$ where Δ is skew-Hermitian.

And now it's clear that, for any $\Delta \in T_I \mathrm{U}(d)$, we get the vector field X_Δ defined by

$$X_\Delta(U) = U\Delta \in T_U \mathrm{U}(d).$$

This is what's called a *left-invariant* vector field on $\mathrm{U}(d)$, since it is invariant under the action of $\mathrm{U}(d)$ on itself by left-multiplication:

$$(dL_U)_V X_\Delta(V) = X_\Delta(UV).$$

Exercise 1.5.6. Check this.

In fact, one can show that all left-invariant vector fields are of this form, so there is an identification between the collection of left-invariant vector fields on $\mathrm{U}(d)$ and $T_I \mathrm{U}(d)$.

- Remark 1.5.7.**
- (i) There was nothing special about $\mathrm{U}(d)$ in Example 1.5.3: we could have used any group G which was a manifold (such groups are called *Lie groups*; more on them later!) and gotten the same correspondence between the tangent space at the identity and left-invariant vector fields. This is important because these are the two standard ways differential geometers think about *Lie algebras*, and this construction shows that they are equivalent.
 - (ii) More generally, the fact that $T_U \mathrm{U}(d)$ consists of matrices of the form $U\Delta$ where Δ is skew-Hermitian means that *every* vector field $X \in \mathfrak{X}(\mathrm{U}(d))$ can be written as

$$X(U) = U\Delta_U,$$

where the skew-Hermitian matrix Δ_U depends on U . So then a vector field on $\mathrm{U}(d)$ induces a mapping $\mathrm{U}(d) \rightarrow T_I \mathrm{U}(d)$ given by $U \mapsto \Delta_U$.

1.5.1 The matrix exponential

This is not directly related to vector fields, but another reason to be interested in $T_I \mathrm{U}(d)$ (or, more generally, the tangent space at the identity of any Lie group) is that it provides local coordinates on almost all of $\mathrm{U}(d)$ by way of the matrix exponential.

First of all, if $\Delta \in T_I \mathrm{U}(d)$ (i.e., Δ is skew-Hermitian), then I claim that $\exp(\Delta) \in \mathrm{U}(d)$, where \exp is the matrix exponential defined by the power series

$$\exp(\Delta) = I + \Delta + \frac{1}{2!}\Delta^2 + \frac{1}{3!}\Delta^3 + \dots$$

In other words, the claim is that $\exp : T_I \mathrm{U}(d) \rightarrow \mathrm{U}(d)$.

To see that $\exp(\Delta) \in \mathrm{U}(d)$, we need to show that $\exp(\Delta) \exp(\Delta)^* = I$, or equivalently $\exp(\Delta)^* = \exp(\Delta)^{-1}$. Now

$$\begin{aligned}\exp(\Delta)^* &= \left(I + \Delta + \frac{1}{2!} \Delta^2 + \frac{1}{3!} \Delta^3 + \dots \right)^* \\ &= I + \Delta^* + \frac{1}{2!} (\Delta^2)^* + \frac{1}{3!} (\Delta^3)^* + \dots \\ &= I + \Delta^* + \frac{1}{2!} (\Delta^*)^2 + \frac{1}{3!} (\Delta^*)^3 + \dots \\ &= I + (-\Delta) + \frac{1}{2!} (-\Delta)^2 + \frac{1}{3!} (-\Delta)^3 + \dots \\ &= \exp(-\Delta),\end{aligned}$$

and by expanding the product of power series one can show that

$$\exp(\Delta) \exp(\Delta)^* = \exp(\Delta) \exp(-\Delta) = \exp(\Delta - \Delta) = I.$$

(It is absolutely essential in the argument for $\exp(\Delta) \exp(-\Delta) = \exp(\Delta - \Delta)$ that Δ and $-\Delta$ commute. When A and B are matrices with $AB \neq BA$, it is not necessarily true that $\exp(A) \exp(B) = \exp(A + B)$; see the Baker–Campbell–Hausdorff formula in general [5, Chapter 5].)

Claim. $\exp: T_I \mathrm{U}(d) \rightarrow \mathrm{U}(d)$ is surjective.

Proof. If $U \in \mathrm{U}(d)$, then the eigenvalues of U are all unit complex numbers $e^{i\theta_1}, \dots, e^{i\theta_d}$, so U has the spectral decomposition

$$U = V \Theta V^*,$$

where

$$\begin{aligned}\Theta &= \mathrm{diag}\left(e^{i\theta_1}, \dots, e^{i\theta_d}\right) \\ &= \mathrm{diag}\left(1 + i\theta_1 + \frac{1}{2!} (i\theta_1)^2 + \dots, \dots, 1 + i\theta_d + \frac{1}{2!} (i\theta_d)^2 + \dots\right) \\ &= \exp(\mathrm{diag}(i\theta_1, \dots, i\theta_d)).\end{aligned}$$

(Note that the set of all such Θ forms a torus; this turns out to be a *maximal torus* inside $\mathrm{U}(d)$).

But then $H = V \mathrm{diag}(i\theta_1, \dots, i\theta_d) V^*$ is skew-Hermitian, and hence in $T_I \mathrm{U}(d)$, and I claim that

$$U = \exp(H).$$

This follows from the more general fact about exponentiating matrix conjugates stated below in [Lemma 1.5.8](#).

Just to verify, let's check this on a random example. Here's a random element of $\mathrm{U}(3)$ (generated in *Mathematica* with `RandomVariate[CircularUnitaryMatrixDistribution[3]]`):

$$U = \begin{bmatrix} -0.392089 + 0.77069i & -0.175305 - 0.0691595i & 0.460085 - 0.0714797i \\ -0.359804 + 0.151554i & 0.618136 + 0.524479i & -0.295949 - 0.32065i \\ 0.103894 + 0.298465i & -0.226492 + 0.505981i & -0.312325 + 0.703749i \end{bmatrix}.$$

Computing the spectral decomposition yields

$$V = \begin{bmatrix} -0.452698 + 0.477605i & 0.724632 + 0.i & -0.203482 + 0.021487i \\ 0.0746037 + 0.313061i & 0.0936111 + 0.271501i & 0.902192 + 0.i \\ 0.680724 + 0.i & 0.346776 + 0.521707i & -0.249272 + 0.286437i \end{bmatrix}$$

and

$$\Theta = \begin{bmatrix} e^{2.58364i} & 0 & 0 \\ 0 & e^{1.68825i} & 0 \\ 0 & 0 & e^{0.496512i} \end{bmatrix}.$$

Therefore,

$$H = \begin{bmatrix} 2.0261i & -0.135698 + 0.322419i & -0.228031 - 0.343709i \\ 0.135698 + 0.322419i & 0.810972i & -0.498786 + 0.313484i \\ 0.228031 - 0.343709i & 0.498786 + 0.313484i & 1.93134i \end{bmatrix},$$

and a calculation shows that $\exp(H) = U$, as desired. \square

Lemma 1.5.8. Suppose $A, B \in \text{Mat}_{d \times d}(\mathbb{C})$. Then

$$\exp(ABA^{-1}) = A \exp(B)A^{-1}.$$

Proof. By definition,

$$\begin{aligned} \exp(ABA^{-1}) &= I + ABA^{-1} + \frac{1}{2!}(ABA^{-1})^2 + \dots = I + ABA^{-1} + \frac{1}{2!}AB^2A^{-1} + \dots \\ &= A \left(I + B + \frac{1}{2!}B^2 + \dots \right) A^{-1} = A \exp(B)A^{-1}. \end{aligned}$$

\square

To recap, we have a surjective map $\exp: T_I \text{U}(d) \rightarrow \text{U}(d)$. Moreover, the non-injectivity of \exp is due to the periodicity of the complex exponential: $e^{i\theta} = e^{i(\theta+2\pi)}$. So \exp maps the neighborhood of the origin in $T_I \text{U}(d)$ consisting of skew-Hermitian matrices with spectral norm $< \pi$ bijectively onto the neighborhood of I in $\text{U}(d)$ consisting of unitary matrices whose eigenvalues all have argument strictly between $-\pi$ and π ; in other words, this only excludes those unitary matrices with i as an eigenvalue, which is some positive codimension subset of $\text{U}(d)$. So then \exp provides a local coordinate chart in the complement of this subset.

1.6 The Lie Bracket on Vector Fields

As we've defined it, $\mathfrak{X}(M)$ is a $C^\infty(M)$ -module:⁶ we can certainly add two vector fields or multiply a vector field by a smooth function and get a new vector field.

⁶If you haven't seen modules before, they're basically like vector fields where the ring of scalars isn't necessarily a field. In the case of $\mathfrak{X}(M)$, the smooth functions $C^\infty(M)$ serve as the scalars, but $C^\infty(M)$ is only a ring, not a field: only nowhere-vanishing functions have well-defined inverses.

In fact, there is even more algebraic structure on $\mathfrak{X}(M)$: it is something called a *Lie algebra*, which means it is a (real) vector space which admits a binary operation (called a *Lie bracket*) satisfying certain axioms.⁷

In order to define the Lie bracket of vector fields, it is instructive to think about how we could possibly define a binary operation on vector fields. For example, we could try to generalize binary operations on vector fields we already know in certain special cases. For example:

- Since $T\mathbb{R} \cong \mathbb{R} \times \mathbb{R}$, we can interpret a vector field on \mathbb{R} as a real-valued function by just recording the second entry: $X(p) = (p, v)$, and the corresponding function is $p \mapsto v$. Since functions form an algebra by pointwise multiplication, this defines a binary operation on vector fields on \mathbb{R} .
- More generally, every smooth 1-manifold has a trivial tangent bundle, so we can treat any vector field on any 1-manifold as a smooth, real-valued function and get an algebra structure on vector fields.⁸
- Since $T\mathbb{R}^2 \cong \mathbb{R}^2 \times \mathbb{R}^2$ and since we can define an equivalence $\mathbb{R}^2 \leftrightarrow \mathbb{C}$ by $(x, y) \leftrightarrow x + iy$, we can interpret a vector field on \mathbb{R}^2 (or, more generally, any surface with trivial tangent bundle) as a complex-valued function. Again, functions form an algebra by pointwise multiplication, so this defines a binary operation on vector fields on \mathbb{R}^2 .⁹
- Since $T\mathbb{R}^3 \cong \mathbb{R}^3 \times \mathbb{R}^3$, we can interpret a vector field on \mathbb{R}^3 as a vector-valued function on \mathbb{R}^3 . Then the (pointwise) cross product \times gives a binary operation on vector fields on \mathbb{R}^3 .
- Since $T\mathbb{R}^4 \cong \mathbb{R}^4 \times \mathbb{R}^4$ and since we can define an equivalence $\mathbb{R}^4 \leftrightarrow \mathbb{H}$ between \mathbb{R}^4 and the quaternions by $(t, x, y, z) \leftrightarrow t + ix + jy + kz$, we can interpret vector fields on \mathbb{R}^4 as quaternion-valued functions. Again, functions form an algebra by pointwise multiplication, so this defines a binary operation on vector fields on \mathbb{R}^4 .

This is in some sense a generalization of the cross product on \mathbb{R}^3 : if we represent $(x, y, z), (u, v, w) \in \mathbb{R}^3$ by purely imaginary quaternions $p = ix + jy + kz$ and $q = iu + jv + kw$, then the quaternion product

$$pq = -p \cdot q + p \times q,$$

where $p \cdot q$ is the dot product of the two vectors in \mathbb{R}^3 , and $p \times q$ is the cross product (interpreted as a purely imaginary quaternion).¹⁰

So the natural question is: can we generalize these examples to higher-dimensional Euclidean spaces, and more generally to arbitrary manifolds?

Unfortunately, the answer is basically “no.” As you might have heard, there are no normed division rings over \mathbb{R} besides \mathbb{R} , \mathbb{C} , and \mathbb{H} , and there aren’t even normed division algebras (where we allow multiplication to be non-associative) aside from these examples and the octonions \mathbb{O} , which are 8-dimensional.

Moreover, even if there were more normed division algebras, the operations above depended very strongly on the triviality of the tangent bundle, which is not going to work in more general manifolds.

⁷Pronunciation note: The Norwegian surname “Lie” is typically pronounced the same as the common English or Korean surname “Lee,” rather than like the English word “lie.”

⁸This correspondence between sections of trivial line bundles and functions gives you some reason to believe that the appropriate notion of “function” in algebraic geometry is often a section of a line bundle.

⁹Interpreting vector fields on the plane as complex-valued functions turns out to be quite useful in 2-dimensional fluid mechanics; look up *stream function* and *complex potential* for more.

¹⁰Look up *Clifford algebras* for a vast generalization of this kind of product operation that splits into a scalar part (the $-p \cdot q$) and a vector part (the $p \times q$).

So here's a different approach: recall that we can interpret vector fields on M as operators on $C^\infty(M)$. That is, for $f \in C^\infty(M)$, $Xf \in C^\infty(M)$ as well; this is basically the function recording the directional derivative of f in the direction of X at each point on the manifold.

An obvious thing to guess is that, if $X, Y \in \mathfrak{X}(M)$, then the composition $X \circ Y = XY$ is also a vector field. After all, $(XY)f = X(Yf)$ will be a smooth function, so XY is also an operator on $C^\infty(M)$.

Let's work in local coordinates at a point $p \in M$ to see what this operator looks like when applied to a function f which is differentiable at p .

We know that, in terms of the local coordinate basis $\left\{\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_n}\right\}$, we can write

$$X(p) = \sum_{i=1}^n a_i(p) \frac{\partial}{\partial x_i} \quad \text{and} \quad Y(p) = \sum_{i=1}^n b_i(p) \frac{\partial}{\partial x_i},$$

where the a_i and b_i are smooth functions defined in some neighborhood of p . So then

$$(XY)f = X(Yf) = X \left(\sum_{i=1}^n b_i \frac{\partial f}{\partial x_i} \right) = \sum_{i,j=1}^n a_j \frac{\partial}{\partial x_j} \left(b_i \frac{\partial f}{\partial x_i} \right) = \sum_{i,j=1}^n a_j \left(\frac{\partial b_i}{\partial x_j} \frac{\partial f}{\partial x_i} + b_i \frac{\partial^2 f}{\partial x_j \partial x_i} \right).$$

This should give you pause because it is no longer a first-order differential operator. Let's make this more obvious by rewriting as

$$(XY)f = \left[\sum_{j=1}^n \left(a_j \left(\sum_{i=1}^n \frac{\partial b_i}{\partial x_j} \frac{\partial}{\partial x_i} \right) \right) + \sum_{j=1}^n \left(a_j \left(\sum_{i=1}^n b_i \frac{\partial^2}{\partial x_j \partial x_i} \right) \right) \right] f, \quad (1.1)$$

so the stuff inside the brackets (which is XY) is a differential operator on $C^\infty(M)$, but it's not a vector field; for example, it cannot be written in terms of the local coordinate basis $\left\{\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_n}\right\}$.

Example 1.6.1. If $X = \frac{\partial}{\partial x}$ and $Y = \frac{\partial}{\partial y}$ are the standard coordinate vector fields on \mathbb{R}^2 , then (1.1) reduces to

$$(XY)f = \frac{\partial^2 f}{\partial x \partial y},$$

as you would expect from computing

$$X(Yf) = \frac{\partial}{\partial x} \left(\frac{\partial f}{\partial y} \right) = \frac{\partial^2 f}{\partial x \partial y}.$$

So $XY = \frac{\partial^2}{\partial x \partial y}$. But there's no sensible way to interpret this second-order operator as a vector field.

Example 1.6.2. If $X = r \frac{\partial}{\partial r}$ and $Y = r \frac{\partial}{\partial \theta}$ are the (scaled) radial and rotational fields, then

$$(XY)f = X(Yf) = r \frac{\partial}{\partial r} \left(r \frac{\partial f}{\partial \theta} \right) = r \frac{\partial f}{\partial \theta} + r^2 \frac{\partial^2 f}{\partial r \partial \theta},$$

so

$$XY = r \frac{\partial}{\partial \theta} + r^2 \frac{\partial}{\partial r \partial \theta}.$$

Again, a second-order operator appears.

The problem with (1.1) is the second term, which is a second-order operator. Now notice that if we had done this in the reverse order we would have gotten

$$(YX)f = \left[\sum_{j=1}^n \left(b_j \left(\sum_{i=1}^n \frac{\partial a_i}{\partial x_j} \frac{\partial}{\partial x_i} \right) \right) + \sum_{j=1}^n \left(b_j \left(\sum_{i=1}^n a_i \frac{\partial^2}{\partial x_j \partial x_i} \right) \right) \right] f. \quad (1.2)$$

This doesn't just have the same type of problem as before, it has literally the same problem: because mixed partials commute, the second terms in (1.1) and (1.2) agree. So, by subtracting, we can cancel them and get

$$(XY - YX)f = \left[\sum_{i,j=1}^n \left(\left(a_j \frac{\partial b_i}{\partial x_j} - b_j \frac{\partial a_i}{\partial x_j} \right) \frac{\partial}{\partial x_i} \right) \right] f, \quad (1.3)$$

which is written in terms of the $\left\{ \frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_n} \right\}$ basis. So this really is a vector field.

Definition 1.6.3. If $X, Y \in \mathfrak{X}(M)$, the *Lie bracket* of X and Y is a vector field $[X, Y]$ defined by

$$[X, Y]f := X(Yf) - Y(Xf).$$

The key feature of this operation is that it satisfies the axioms of a Lie bracket ((i)–(iii) in Proposition 1.6.4), and hence makes $\mathfrak{X}(M)$ into a *Lie algebra*.

Proposition 1.6.4. If $X, Y, Z \in \mathfrak{X}(M)$ and $a, b \in \mathbb{R}$, $f, g \in C^\infty(M)$, then

- (i) $[X, Y] = -[Y, X]$ (anti-commutativity)
- (ii) $[aX + bY, Z] = a[X, Z] + b[Y, Z]$ (linearity)
- (iii) $[[X, Y], Z] + [[Y, Z], X] + [[Z, X], Y] = 0$ (Jacobi identity)
- (iv) $[fX, gY] = fg[X, Y] + f(Xg)Y - g(Yf)X$.

Proof. Homework. □

Example 1.6.5. Continuing with Example 1.6.2, we already computed XY , so we can also compute

$$YX = r \frac{\partial}{\partial \theta} \left(r \frac{\partial}{\partial r} \right) = r^2 \frac{\partial^2}{\partial \theta \partial r},$$

and hence

$$[X, Y] = r \frac{\partial}{\partial \theta}.$$

Example 1.6.6. Consider $M = S^2$ and the vector fields $X = \frac{\partial}{\partial \theta}$ and $Y = \frac{\partial}{\partial z}$ where (θ, z) are cylindrical coordinates on S^2 ; see Figure 1.11.

In other words, we have the local coordinate chart $\phi: (0, 2\pi) \times (-1, 1) \rightarrow S^2$ given by

$$\phi(\theta, z) = (\sqrt{1 - z^2} \cos \theta, \sqrt{1 - z^2} \sin \theta, z),$$

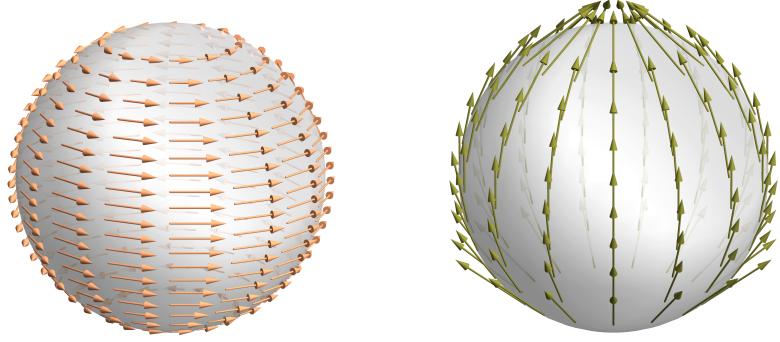


Figure 1.11: The vector fields $X = \frac{\partial}{\partial\theta}$ and $Y = \frac{\partial}{\partial z}$ on S^2 .

and X and Y are the corresponding coordinate fields.

Notice that

$$[X, Y]f = \frac{\partial^2 f}{\partial\theta\partial z} - \frac{\partial^2 f}{\partial z\partial\theta} = 0$$

since mixed partials commute.

More generally, whenever $X = \frac{\partial}{\partial x_i}$ and $Y = \frac{\partial}{\partial x_j}$ are coordinate fields in a neighborhood of a point in a manifold, $[X, Y] = 0$.

In Cartesian coordinates

$$X = \sqrt{1-z^2} \left(-\sin\theta \frac{\partial}{\partial x} + \cos\theta \frac{\partial}{\partial y} \right) = \sqrt{1-z^2} \left(-y \frac{\partial}{\partial x} + x \frac{\partial}{\partial y} \right),$$

and

$$Y = -\frac{z \cos\theta}{\sqrt{1-z^2}} \frac{\partial}{\partial x} - \frac{z \sin\theta}{\sqrt{1-z^2}} \frac{\partial}{\partial y} + \frac{\partial}{\partial z} = -\frac{xz}{1-z^2} \frac{\partial}{\partial x} - \frac{yz}{1-z^2} \frac{\partial}{\partial y} + \frac{\partial}{\partial z}.$$

So then some much more unpleasant calculations using (1.3) shows you that $[X, Y] = 0$ in these coordinates as well.

Now we look at some examples of Lie algebras not coming from vector fields.

Example 1.6.7. \mathbb{R}^3 forms a Lie algebra with the bracket operation given by the cross product: for $u, v \in \mathbb{R}^3$, define $[u, v] := u \times v$. Then it is straightforward to check that this bracket satisfies (i)–(iii) above. Written in terms of cross products, (iii) says

$$0 = (u \times v) \times w + (v \times w) \times u + (w \times u) \times v = (u \times v) \times w - u \times (v \times w) + v \times (u \times w),$$

or equivalently,

$$(u \times v) \times w = u \times (v \times w) - v \times (u \times w).$$

In particular, this records the failure of associativity of the cross product.

More generally, the Jacobi identity records the failure of associativity of the Lie bracket:

$$[[X, Y], Z] = [X, [Y, Z]] - [Y, [X, Z]].$$

Example 1.6.8. Recall [Example 1.5.3](#), in which we showed that left-invariant vector fields on $U(d)$ are of the form $X(U) = U\Delta_X$, for $\Delta_X \in T_I U(d)$, which is the collection of skew-Hermitian $d \times d$ matrices. In particular, $X(I) = \Delta_X$.

As we will see in more detail later ([Proposition 3.4.2](#)), it will turn out that the Lie bracket on left-invariant vector fields on a matrix group like $U(d)$ just corresponds to the matrix commutator operation in the tangent space to the identity.

In this case, that means that, if $X, Y \in \mathfrak{X}(U(d))$ are left-invariant, then $[X, Y]$ is also left-invariant and, for each $U \in U(d)$,

$$[X, Y](U) = U(\Delta_X \Delta_Y - \Delta_Y \Delta_X),$$

where the products inside parentheses are just matrix products between the skew-Hermitian matrices Δ_X and Δ_Y ; that is, the term in parentheses is just the usual matrix commutator.

Exercise 1.6.9. Check that the commutator of two skew-Hermitian matrices is skew-Hermitian.

This all tells you that the correspondence between left-invariant vector fields and elements of $T_I U(d)$ turns the Lie bracket of left-invariant vector fields into the matrix commutator in $T_I U(d)$. These are two different realizations of the same Lie algebra, usually called $\mathfrak{u}(d)$.

Example 1.6.10. If we play the same game with $SO(d)$, it turns out that the tangent space to the identity consists of $d \times d$ skew-symmetric matrices, and the Lie algebra on left-invariant vector fields on $SO(d)$ corresponds to the matrix commutator on skew-symmetric matrices; we'll write the collection of skew-symmetric $d \times d$ matrices as $\mathfrak{so}(d)$.

Consider the case $d = 3$. Then we can write elements of $\mathfrak{so}(3)$ as

$$\Delta = \begin{bmatrix} 0 & -z & y \\ z & 0 & -x \\ -y & x & 0 \end{bmatrix}.$$

(The reason for the funny ordering and sign choices will shortly become apparent.)

Now, $\mathfrak{so}(3)$ is a 3-dimensional vector space, and we have a vector space isomorphism $F: \mathbb{R}^3 \rightarrow \mathfrak{so}(3)$ given by

$$F: (x, y, z) \mapsto \begin{bmatrix} 0 & -z & y \\ z & 0 & -x \\ -y & x & 0 \end{bmatrix}. \quad (1.4)$$

Let $\Delta_1, \Delta_2 \in \mathfrak{so}(3)$ with

$$\Delta_i = \begin{bmatrix} 0 & -z_i & y_i \\ z_i & 0 & -x_i \\ -y_i & x_i & 0 \end{bmatrix}.$$

Then

$$[\Delta_1, \Delta_2] = \Delta_1 \Delta_2 - \Delta_2 \Delta_1 = \begin{bmatrix} 0 & x_2 y_1 - x_1 y_2 & x_2 z_1 - x_1 z_2 \\ x_1 y_2 - x_2 y_1 & 0 & y_2 z_1 - y_1 z_2 \\ x_1 z_2 - x_2 z_1 & y_1 z_2 - y_2 z_1 & 0 \end{bmatrix}.$$

If you stare at this, you might recognize the entries as being the coordinates of the cross product of the corresponding vectors:

$$\begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} \times \begin{bmatrix} x_2 \\ y_2 \\ z_2 \end{bmatrix} = \begin{bmatrix} y_1 z_2 - y_2 z_1 \\ x_2 z_1 - x_1 z_2 \\ x_1 y_2 - x_2 y_1 \end{bmatrix}.$$

In other words, $F(v_1 \times v_2) = [F(v_1), F(v_2)]$, so F is a Lie algebra homomorphism. Since it's also a bijective linear map, it's a Lie algebra isomorphism, so we've just proved that $(\mathbb{R}^3, \times) \cong \mathfrak{so}(3)$ as Lie algebras.

Thus, the Lie bracket on vector fields on manifolds is some sort of vast generalization of the cross product on \mathbb{R}^3 . In this interpretation, $\mathbb{R}^3 \cong \mathfrak{so}(3)$ is the collection of infinitesimal rotations of 3-space, where $v \in \mathbb{R}^3$ corresponds to an infinitesimal rotation around the axis spanned by v , and the correspondence between cross products and commutators reflects the fact that, for very small $\epsilon > 0$ and unit vectors u and v , the composition of an ϵ -rotation around u with an ϵ -rotation around v is, to first order, a rotation around $u + v + \frac{\epsilon}{2}u \times v$; see the Baker–Campbell–Hausdorff formula [5, Chapter 5].

Exercise 1.6.11. Convince yourself that, for $v \in \mathbb{R}^3$ a unit vector and F as defined in (1.4), $\exp(F(\theta v))$ gives the one-parameter subgroup of rotations by angle θ around the axis spanned by v . (A full proof is kind of annoying to write down, but at least convince yourself this is true when v is a coordinate vector.)

1.7 Integral Curves and Lie Derivatives

To motivate the definition of the Lie bracket, I claimed that there was additional algebraic structure on $\mathfrak{X}(M)$ in the form of a binary operation, and then (hopefully!) convinced you that there was really only one sensible way to define such an operation.

But it's probably not at all obvious why one should have guessed that there was a binary operation on vector fields in the first place, and, though I don't actually know the history, my guess is that this was not the original motivation for the Lie bracket.

In some sense the more basic notion is that of the *Lie derivative*, which is a sort of directional derivative of vector fields which, as we will see, actually generalizes to arbitrary tensor fields.

The idea is that, given two vector fields X and Y on a manifold M , we might like to define a “directional derivative of Y in the direction of X ” operator, which we will denote $\mathcal{L}_X Y$, where the \mathcal{L} is for “Lie derivative.” It will turn out that $\mathcal{L}_X Y = [X, Y]$, which might be surprising, but if you look back to the local coordinate expression (1.3) for $[X, Y]$, you'll notice it involved differentiating the coefficients of Y with respect to X (and also X with respect to Y , but we'll shortly see why there's this symmetry).

It's worth thinking for yourself about how you might go about defining such a derivative operator, so I would encourage you to do so before reading on.

Here's a way of making this “directional derivative for vector fields” operation more precise: say we have our vector fields X and Y and we want to compute $\mathcal{L}_X Y$ at a point $p \in M$. Presumably the idea is to measure how much Y is changing as we vary p in the direction of X .

In other words, we want to integrate X to get a curve $\alpha : (-\epsilon, \epsilon) \rightarrow M$ so that $\alpha(0) = p$ and $\alpha'(t) = X(\alpha(t))$ for all $t \in (-\epsilon, \epsilon)$, then look at something like

$$\lim_{t \rightarrow 0} \frac{Y(\alpha(t)) - Y(\alpha(0))}{t}.$$

See Figure 1.12.



Figure 1.12: An attempt to differentiate Y in the direction of X .

In other words, we look at the rate of change of Y as we flow in the direction of X . Unfortunately, as written the above difference quotient doesn't make sense, but first let's talk about flows and integrating vector fields.

Definition 1.7.1. Let $X \in \mathfrak{X}(M)$. A curve $\alpha: (a, b) \rightarrow M$ is called an *integral curve* (or *trajectory*) for X if $\alpha'(t) = X(\alpha(t))$ for all $t \in (a, b)$.

Proposition 1.7.2. Let $X \in \mathfrak{X}(M)$ and let $p \in M$. Then there exists a neighborhood U of p , a $\delta > 0$, and a smooth map $\Phi: (-\delta, \delta) \times U \rightarrow M$ so that, for each $q \in U$, the map $t \mapsto \Phi(t, q)$ is the unique smooth curve satisfying the ODE $\frac{\partial \Phi}{\partial t} = X(\Phi(t, q))$ with initial condition $\Phi(0, q) = q$. See [Figure 1.13](#).

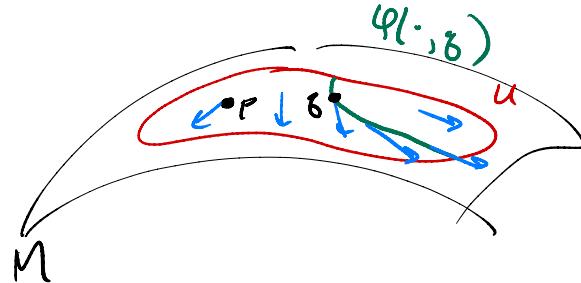


Figure 1.13: The local flow generates an integral curve through each point.

Proof. Since M is locally diffeomorphic to \mathbb{R}^n (via the coordinate charts), this follows from the fundamental theorem on existence and uniqueness of solutions to ODEs in \mathbb{R}^n . \square

We often use the notation $\Phi_t(q) := \Phi(t, q)$ and call Φ_t the *local flow* of X .

Example 1.7.3. Let $M = S^2$ and consider the vector field $X(x, y, z) = zx \frac{\partial}{\partial x} + zy \frac{\partial}{\partial y} + (z^2 - 1) \frac{\partial}{\partial z}$ (of course, this is written in extrinsic \mathbb{R}^3 coordinates rather than intrinsic coordinates). See [Figure 1.14](#).

In fact, you can check that this is just $(d\phi_N)Y$, where Y is the vector field on \mathbb{R}^2 given by $Y(u, v) = -u \frac{\partial}{\partial u} - v \frac{\partial}{\partial v}$.

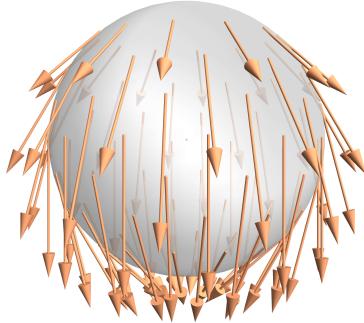


Figure 1.14: The vector field $X(x, y, z) = zx \frac{\partial}{\partial x} + zy \frac{\partial}{\partial y} + (z^2 - 1) \frac{\partial}{\partial z}$ on the sphere.

Visually, the local flow pushes the mass of the sphere down towards the south pole. We can integrate the flow explicitly in *Mathematica* to get

$$\Phi_t(x, y, z) = \frac{1}{1 + 2e^{2t} + (1 - e^{2t})z} \left(2e^t x, 2e^t y, 1 - e^{2t} + (1 + e^{2t})z \right).$$

See [Figure 1.15](#).

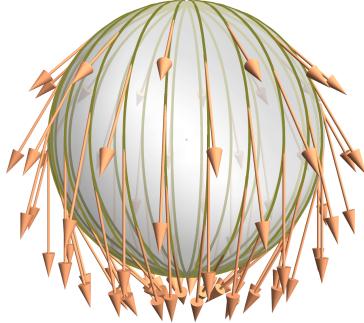


Figure 1.15: The local flow of X .

This is an example of a *complete vector field* since the local flow exists for all time at all points. If we had removed the Antarctic Circle from the sphere, this vector field would be incomplete (since the flow ceases to exist once you go over the edge).

With local flows in our toolkit we would then like to define the Lie derivative as something like

$$(\mathcal{L}_X Y)(p) = \lim_{t \rightarrow 0} \frac{Y(\Phi_t(p)) - Y(p)}{t}.$$

But if you think about it, the numerator in the difference quotient *makes no sense!* After all, $Y(\Phi_t(p)) \in T_{\Phi_t(p)}M$ and $Y(p) \in T_pM$ live in completely different vector spaces. There's no sense in which we can subtract these two vectors.

So we need some way to get the tangent vectors determined by Y at different points to be in the same tangent space.

Thus far, the only machines we have for moving tangent vectors between different tangent spaces are differentials of smooth maps. The only smooth map we have at our disposal is Φ_t , so somehow that must come into play. One possibility is

$$\lim_{t \rightarrow 0} \frac{Y(\Phi_t(p)) - (d\Phi_t)_p Y(p)}{t}$$

since both $Y(\Phi_t(p))$ and $(d\Phi_t)_p Y(p)$ live in $T_{\Phi_t(p)}M$. The problem with this is that, as t changes, $T_{\Phi_t(p)}M$ is also changing, so we're taking limits of vectors in different tangent spaces. Now, all these tangent vectors still live in TM , so this is in fact doable, but it would be better if all the vectors in the limit actually lived in the tangent space, ideally T_pM .

We are finally ready to define the Lie derivative:

Definition 1.7.4. Suppose M is a manifold, $X, Y \in \mathfrak{X}(M)$, and $p \in M$. Then the *Lie derivative* of Y with respect to X at p is

$$(\mathcal{L}_X Y)(p) := \lim_{t \rightarrow 0} \frac{(d\Phi_{-t})_{\Phi_t(p)} Y(\Phi_t(p)) - Y(p)}{t},$$

where Φ_t is the local flow for X .

This is a lot of obnoxious notation, but let's try to unpack it. First of all, notice that $\Phi_{-t}(\Phi_t(p)) = p$, since we're just flowing forward in time by t and then backwards for the same time. So

$$(d\Phi_{-t})_{\Phi_t(p)} : T_{\Phi_t(p)}M \rightarrow T_pM.$$

Hence, plugging in $Y(\Phi_t(p))$ gives $(d\Phi_{-t})_{\Phi_t(p)} Y(\Phi_t(p)) \in T_pM$, and the difference quotient and the limit make sense as happening entirely in T_pM . A (somewhat messy) picture is shown in [Figure 1.16](#).



Figure 1.16: The pieces in the definition of $\mathcal{L}_X Y$.

In looking at [Figure 1.16](#), recall that the point was that we wanted to see how Y varied as we moved in the direction of X . So at a point $p \in M$, we follow the local flow of X for a small time t . Y determines a tangent vector at the resulting point $\Phi_t(p)$. Now we push this vector $Y(\Phi_t(p))$ to T_pM by the negative local flow of X and compare to $Y(p)$.

You can imagine that if Y had a constant magnitude and made a constant angle with X , then pushing Y forward by the negative local flow (which is exactly $(d\Phi_{-t})_{\Phi_t(p)}Y(\Phi_t(p))$) would just give you the same vector you started with, namely $Y(p)$. But this makes sense: if Y has a constant magnitude and makes a constant angle with X , then it's not changing at all with respect to X , and this derivative should really be 0.¹¹

The slightly amazing fact is that the Lie derivative and the Lie bracket are the same:

Proposition 1.7.5. *Let $X, Y \in \mathfrak{X}(M)$ and choose $p \in M$. Then*

$$[X, Y](p) = (\mathcal{L}_X Y)(p).$$

Notice, in particular, that this shows that the Lie derivative is antisymmetric:

$$\mathcal{L}_Y X = [Y, X] = -[X, Y] = -\mathcal{L}_X Y$$

by the antisymmetry of the Lie bracket, which you might not have guessed from the definition of the Lie derivative.

Proof of Proposition 1.7.5. Both expressions are tangent vectors at p , so formally they're differential operators evaluated on functions that are differentiable in a neighborhood of p . So we'll choose a test function f which is differentiable in a neighborhood of p , and the goal is to show that

$$\lim_{t \rightarrow 0} \frac{(d\Phi_{-t})_{\Phi_t(p)}Y(\Phi_t(p)) - Y(p)}{t}(f)(p) = ((XY - YX)f)(p).$$

Let's focus on the first term in the numerator, $((d\Phi_{-t})_{\Phi_t(p)}Y)f(p)$:

$$((d\Phi_{-t})_{\Phi_t(p)}Y)f(p) = (df_p)((d\Phi_{-t})_{\Phi_t(p)}Y(p)) = (d(f \circ \Phi_{-t}))_{\Phi_t(p)}Y(\Phi_t(p)) = (Y(f \circ \Phi_{-t}))(\Phi_t(p)),$$

where we used Lemma 1.3.4 for the first and last equalities¹² and the chain rule for the middle equality. Therefore,

$$(\mathcal{L}_X Y)(p) = \lim_{t \rightarrow 0} \frac{(Y(f \circ \Phi_{-t}))(\Phi_t(p)) - (Yf)(p)}{t}.$$

If we could turn this into something like

$$\lim_{t \rightarrow 0} \frac{(Yf)(\Phi_t(p)) - (Yf)(p)}{t} = \frac{d((Yf) \circ \Phi_t)}{dt} \Big|_{t=0} = (X(Yf))(p), \quad (1.5)$$

we'd be in business.

In fact, if we could write $f(\Phi_{-t}(q)) = f(q) + tg(t, q)$ for some function g , then we could get a term

¹¹You should take this paragraph figuratively and not literally: we don't have any notion of magnitude or angle for tangent vectors, because we don't yet have inner products on the tangent spaces. That will come when we start talking about Riemannian metrics!

¹²Note: I added this lemma after the initial draft, so it was not there when you previously read that section.

like (1.5) to appear because then we'd have

$$\begin{aligned}
\lim_{t \rightarrow 0} \frac{(Y(f \circ \Phi_{-t}))(\Phi_t(p)) - (Yf)(p)}{t} &= \lim_{t \rightarrow 0} \frac{(Y(f + tg(t, \cdot)))(\Phi_t(p)) - (Yf)(p)}{t} \\
&= \lim_{t \rightarrow 0} \frac{(Yf)(\Phi_t(p)) + (Y(tg(t, \cdot)))(\Phi_t(p)) - (Yf)(p)}{t} \\
&= \lim_{t \rightarrow 0} \frac{(Yf)(\Phi_t(p)) + t(Y(g(t, \cdot)))(\Phi_t(p)) - (Yf)(p)}{t} \\
&= \lim_{t \rightarrow 0} \frac{(Yf)(\Phi_t(p)) - (Yf)(p)}{t} + Y(g(0, \Phi_0(p))) \\
&= (X(Yf))(p) + Y(g(0, p)).
\end{aligned}$$

This will then equal $((XY - YX)f)(p)$ if we can find such a g so that the expression $Y(g(0, p))$ is equal to $(Y(Xf))(p)$.

But this is easy: rearranging $f(\Phi_{-t}(q)) = f(q) + tg(t, q)$ yields

$$g(t, q) = \frac{f(\Phi_{-t}(q)) - f(q)}{t},$$

at least for $t \neq 0$. We can extend the definition to $t = 0$ by taking the limit:

$$g(0, q) := \lim_{t \rightarrow 0} \frac{f(\Phi_{-t}(q)) - f(q)}{t} = (Xf)(q),$$

which is exactly what we need, since then $Y(g(0, p)) = (Y(Xf))(p)$. \square

1.8 Extended Example on S^3

Let S^3 be the unit sphere in $\mathbb{R}^4 \cong \mathbb{C}^2 \cong \mathbb{H}$, where

$$\mathbb{H} = \{a + bi + cj + dk : a, b, c, d \in \mathbb{R}\}$$

is the division ring of quaternions. We have the defining conditions

$$i^2 = j^2 = k^2 = -1, \quad ijk = -1,$$

which for example imply that

$$-ij = ijk^2 = (ijk)k = -k \Rightarrow ij = k$$

and

$$-ji = ji(ijk) = ji^2jk = -j^2k = k \Rightarrow ji = -k,$$

so \mathbb{H} is noncommutative. \mathbb{H} also has a conjugation operation defined by

$$\overline{a + bi + cj + dk} = a - bi - cj - dk$$

and an inner product given by,

$$\langle p, q \rangle = \operatorname{Re}(p\bar{q}).$$

When $p = a + bi + cj + dk$ and $q = t + xi + yj + zk$, we see that

$$\langle p, q \rangle = \operatorname{Re}(p\bar{q}) = \operatorname{Re}((a + bi + cj + dk)(t - xi - yj - zk)) = at + bx + cy + dz,$$

so this agrees with the usual dot product on \mathbb{R}^4 . Moreover, the induced norm

$$\|p\|^2 = \langle p, p \rangle = \operatorname{Re}(p\bar{p}) = \operatorname{Re}((a + bi + cj + dk)(a - bi - cj - dk)) = a^2 + b^2 + c^2 + d^2$$

is the standard one.

Thought of as the unit quaternions $\{q \in \mathbb{H} : \|q\| = 1\}$, it becomes clear that S^3 is a group: if $p = a + bi + cj + dk, q = t + xi + yj + zk \in \mathbb{H}$ are both unit quaternions, meaning that

$$a^2 + b^2 + c^2 + d^2 = p\bar{p} = \|p\|^2 = 1 = \|q\|^2 = q\bar{q} = t^2 + x^2 + y^2 + z^2,$$

then their product is also a unit quaternion:

$$\|pq\| = (pq)(\bar{pq}) = pq\bar{q}\bar{p} = p1\bar{p} = p\bar{p} = 1.$$

To give some other terminology, S^3 is the *symplectic group* $\operatorname{Sp}(1)$; that is, the quaternionic analog of the unitary group for \mathbb{H}^1 . (We will eventually see that this group is also isomorphic to $\operatorname{SU}(2)$ and $\operatorname{Spin}(3)$).

Geometrically, it is clear that, for any $p \in S^3$, the tangent space $T_p S^3$ can be thought of as the quaternions which are orthogonal to p . So for example the tangent space to the identity $1 \in S^3$ is

$$T_1 S^3 = \{xi + yj + zk : x, y, z \in \mathbb{R}^3\},$$

the purely imaginary quaternions. Using the same idea as in [Example 1.5.3](#), we can push any tangent vector at the identity around to get a left-invariant vector field on all of S^3 . In particular, we get three mutually perpendicular vector fields $X, Y, Z \in \mathfrak{X}(S^3)$ given by

$$X(p) = pi, \quad Y(p) = pj, \quad Z(p) = pk$$

for all $p \in S^3$. (Strictly speaking, $X(p) = (dL_p)_1 i$, where L_p is left-multiplication by p , and i is interpreted as an element of $T_1 S^3$ [and similarly for Y and Z], but L_p being linear implies that $(dL_p)_1 i = pi$.)

To check that, say, $X(p)$ is really a tangent vector at p , notice that

$$\langle p, X(p) \rangle = \langle p, pi \rangle = \langle a + bi + cj + dk, -b + ai + dj - ck \rangle = -ab + ab - cd + cd = 0,$$

so $X(p) \perp p$ and hence $X(p) \in T_p S^3$. Similar calculations show that $Y(p), Z(p) \in T_p S^3$.

Moreover X, Y , and Z are mutually perpendicular everywhere: for example

$$\langle X(p), Y(p) \rangle = \langle pi, pj \rangle = \langle -b + ai + dj - ck, -c - di + aj + bk \rangle = bc - ad + ad - bc = 0.$$

(A better argument: multiplication by a unit quaternion is an isometry of $\mathbb{H} = \mathbb{R}^4$, so orthogonality of $X(1) = i$ and $Y(1) = j$ carries over to $T_p S^3$.)

Now, the corresponding local flows are

$$\xi_t(p) = p(\cos t + i \sin t), \quad \psi_t(p) = p(\cos t + j \sin t), \quad \zeta_t(p) = p(\cos t + k \sin t),$$

as we can see by differentiating; for example,

$$\frac{d}{dt} \Big|_{t=0} \xi_t(p) = \frac{d}{dt} \Big|_{t=0} p(\cos t + i \sin t) = -p \sin(0) + pi \cos(0) = pi = X(p).$$

Now we compute some Lie brackets. For example,

$$\begin{aligned} [X, Y](p) &= (\mathcal{L}_X Y)(p) = \lim_{t \rightarrow 0} \frac{(d\xi_{-t})_{\xi_t(p)} Y(\xi_t(p)) - Y(p)}{t} \\ &= \frac{d}{dt} \Big|_{t=0} (d\xi_{-t})(Y(\xi_t(p))) \\ &= \frac{d}{dt} \Big|_{t=0} (d\xi_{-t})(\xi_t(p)j) \\ &= \frac{d}{dt} \Big|_{t=0} (d\xi_{-t})(p(\cos t + i \sin t)j) \\ &= \frac{d}{dt} \Big|_{t=0} (d\xi_{-t})(p(j \cos t + k \sin t)). \end{aligned}$$

The differential $(d\xi_{-t})_{\xi_t(p)} : T_{\xi_t(p)} S^3 \rightarrow T_p S^3$, but we can interpret both of these as subspaces of the ambient space \mathbb{R}^4 , so the differential can be represented by a 4×4 matrix, namely

$$(d\xi_{-t})_{\xi_t(p)} = \begin{bmatrix} \cos t & \sin t & 0 & 0 \\ -\sin t & \cos t & 0 & 0 \\ 0 & 0 & \cos t & -\sin t \\ 0 & 0 & \sin t & \cos t \end{bmatrix},$$

so

$$(d\xi_{-t})(p(j \cos t + k \sin t)) = \begin{bmatrix} \cos t & \sin t & 0 & 0 \\ -\sin t & \cos t & 0 & 0 \\ 0 & 0 & \cos t & -\sin t \\ 0 & 0 & \sin t & \cos t \end{bmatrix} \begin{bmatrix} -c \cos t - d \sin t \\ c \sin t - d \cos t \\ a \cos t - b \sin t \\ a \sin t + b \cos t \end{bmatrix} = \begin{bmatrix} -c \cos 2t - d \sin 2t \\ c \sin 2t - d \cos 2t \\ a \cos 2t - b \sin 2t \\ a \sin 2t + b \cos 2t \end{bmatrix}.$$

Putting this all together, then

$$[X, Y](p) = \frac{d}{dt} \Big|_{t=0} (d\xi_{-t})(p(j \cos t + k \sin t)) = \begin{bmatrix} -2d \\ 2c \\ -2b \\ 2a \end{bmatrix} = 2pk = 2Z(p).$$

In other words, $[X, Y] = 2Z$ and, by similar arguments, $[Y, Z] = 2X$ and $[Z, X] = 2Y$.

In our language from [Example 1.5.3](#), X , Y , and Z are left-invariant vector fields on the Lie group S^3 , and so you should expect that the Lie bracket on left-invariant vector fields on S^3 corresponds to a Lie algebra structure on $T_1 S^3$, which is a 3-dimensional vector space. We'll call this 3-dimensional Lie algebra $\mathfrak{sp}(1)$ since $S^3 = \text{Sp}(1)$.

Thinking of S^3 and $\mathfrak{sp}(1)$ as living in the space of 1×1 quaternionic matrices (that is, inside \mathbb{H}), we might expect that the matrix commutator on $\mathfrak{sp}(1) = \{xi + yj + zk\}$ should correspond to the above Lie

bracket. Indeed, the commutator of the 1×1 matrices i and j is

$$ij - ji = k - (-k) = 2k,$$

which agrees with the above calculation that $[X, Y] = 2Z$.

Since the only Lie algebra structure we've seen on a 3-dimensional vector space is that of (\mathbb{R}^3, \times) , or equivalently $(\mathfrak{so}(3), [\cdot, \cdot])$, you might guess that $\mathfrak{sp}(1)$ is just another iteration of the same Lie algebra.

Indeed, if e_1, e_2, e_3 is the standard basis for \mathbb{R}^3 and we define $F: \mathbb{R}^3 \rightarrow \mathfrak{sp}(1)$ by

$$F(e_1) = \frac{1}{2}X, \quad F(e_2) = \frac{1}{2}Y, \quad F(e_3) = \frac{1}{2}Z,$$

then we see that, e.g.,

$$F(e_1 \times e_2) = F(e_3) = \frac{1}{2}Z = \frac{1}{4}[X, Y] = \left[\frac{1}{2}X, \frac{1}{2}Y \right] = [F(e_1), F(e_2)],$$

and more generally it's easy to check that $F(u \times v) = [F(u), F(v)]$ for any $u, v \in \mathbb{R}^3$, so F is an isomorphism of Lie algebras.

This also implies that $\mathfrak{so}(3)$ and $\mathfrak{sp}(1)$ are isomorphic Lie algebras, even though the Lie groups $\text{SO}(3)$ and S^3 are not even homotopy equivalent (for example, $\pi_1(\text{SO}(3)) \cong \mathbb{Z}/2\mathbb{Z}$ and $\pi_1(S^3) = \{1\}$), which is going to imply that they cannot be isomorphic as Lie groups.¹³

¹³That said, $\text{SO}(3) \cong \mathbb{RP}^3$ has S^3 as its universal cover; indeed, this is really what the isomorphism of Lie algebras implies.

Chapter 2

Differential Forms

2.1 Integration, Forms, and an Informal Definition

Now that we've defined vector fields and explored some special features (like the Lie bracket), the next objects to turn our attention to are *differential forms*. Differential forms turn out to be the right tool with which to define integration on manifolds, and they also turn out to encode the cohomology ring of a manifold, which means they have both a natural binary operation (corresponding to the cup product in cohomology) and a natural derivative operation (corresponding to the coboundary map).

To try to build up some intuition for differential forms, think back to vector calculus and computing integrals of the form

$$\int_A f \, dx_1 \dots dx_n, \tag{2.1}$$

where $A \subseteq \mathbb{R}^n$ is an open set (or maybe the closure of an open set), and $f : A \rightarrow \mathbb{R}$ is a (sufficiently nice) function.¹ If you stare at (2.1), and in particular at the expression $f \, dx_1 \dots dx_n$ being integrated, what exactly is this thing?

To get a handle on what it is, it's helpful to see how it transforms. If $B \subseteq \mathbb{R}^n$ is open and $g : B \rightarrow A$ is a diffeomorphism,² then

$$\int_A f \, dx_1 \dots dx_n = \int_B (f \circ g) |\det Jg| \, dy_1 \dots dy_n,$$

where I'm using y_1, \dots, y_n to indicate the coordinates on B and Jg is the Jacobian matrix of g . So now whatever kind of object $f \, dx_1 \dots dx_n$ is, it's the same type of object as $(f \circ g) |\det Jg| \, dy_1 \dots dy_n$, and this describes how it transforms under diffeomorphisms.

Now, examining $(f \circ g) |\det Jg| \, dy_1 \dots dy_n$, it's clear that this is a more complicated object than just a function. In particular, the presence of the determinant tells you that this is some sort of *alternating, multilinear* object. Recall the definitions:

Definition 2.1.1. Let V_1, \dots, V_k, W be R -modules. A map

$$f : V_1 \times \dots \times V_k \rightarrow W$$

¹For our purposes we'll just be dealing with the Riemann integral, but one can also generalize the Lebesgue integral to manifolds.

²In calculus terms, this is a change of variables; e.g., a transition from Cartesian to spherical coordinates.

is *multilinear* if it is linear in each factor. That is, for any $j \in \{1, \dots, k\}$, if $v_i \in V_i$ for each $i = 1, \dots, k$, $u_j \in V_j$, and $c \in \mathbb{R}$, then

$$f(v_1, \dots, u_j + cv_j, \dots, v_k) = f(v_1, \dots, u_j, \dots, v_k) + cf(v_1, \dots, v_j, \dots, v_k).$$

Definition 2.1.2. Let V and W be R -modules. A multilinear map $f: V^k \rightarrow W$ is *alternating* if, for any $i \neq j$, $v_i = v_j$ implies that $f(v_1, \dots, v_k) = 0$.

Thinking of the determinant map as a function $\det: V^n \rightarrow \mathbb{R}$ where $(v_1, \dots, v_n) \in V^n$ are interpreted as the columns of a matrix whose determinant is then computed, \det is multilinear and alternating. Indeed, these two properties uniquely characterize the determinant up to scale:

Theorem 2.1.3. Suppose V is an n -dimensional vector space. Up to scaling by a constant factor, \det is the unique alternating, multilinear map $V^n \rightarrow \mathbb{R}$.

The point of all of this is that the things we actually integrate in multivariable calculus are alternating, multilinear gadgets. Informally, this is how we should think about differential forms on manifolds:

Definition 2.1.4 (Informal Definition). A smooth *differential k -form* on a manifold M is a smooth, alternating, $C^\infty(M)$ -multilinear map

$$\omega: \mathfrak{X}(M)^k \rightarrow C^\infty(M)$$

(recall that $\mathfrak{X}(M)$ is a $C^\infty(M)$ -module). The vector space of k -forms is denoted $\Omega^k(M)$.

In other words, at each point $p \in M$, a k -form ω will input k tangent vectors at p and output a real number in an alternating, multilinear way. You might guess that this means that ω is *really* a (smooth) section of some vector bundle over M , and indeed we will shortly define it in this way.

But before that, let's look at some examples of things that should be differential forms according to whatever formal definition we eventually give.

Example 2.1.5. Let $k = 0$. What is a 0-form? Just from looking at [Definition 2.1.4](#), it's supposed to be something which inputs 0 vector fields on M and outputs a smooth function on M . But then that means it has no input and a smooth function as output, so is really just that smooth function. So $C^\infty(M) = \Omega^0(M)$.

Example 2.1.6. If $k > \dim(M)$, then the alternating condition implies that $\Omega^k(M) = \{0\}$.

Example 2.1.7. Suppose $f \in C^\infty(M)$; that is, $f: M \rightarrow \mathbb{R}$ is smooth. Recall that, for each $p \in M$, the differential $df_p: T_p M \rightarrow T_{f(p)} \mathbb{R}$. But $T_{f(p)} \mathbb{R}$ can be canonically identified with \mathbb{R} , so we can interpret df_p as a linear map $T_p M \rightarrow \mathbb{R}$.

In plainer language, df inputs a vector field on M and outputs a number at each point—that is, a function on M . So in fact, since the alternating condition is vacuous with only one input, $df \in \Omega^1(M)$.

(Foreshadowing: $f \in \Omega^0(M)$ and $df \in \Omega^1(M)$, so you might ask whether in general there's some operation d which turns elements of $\Omega^k(M)$ into elements of $\Omega^{k+1}(M)$.)

Example 2.1.8. Suppose $\alpha \in \Omega^1(M)$. Then at each point α inputs a tangent vector and outputs a number in a linear way. In other words, for each $p \in M$, $\alpha_p \in (T_p M)^*$, the dual space of $T_p M$, also called the cotangent space at p .

Just as a vector field is a (smooth) choice of tangent vector at each point, this says that a 1-form is a (smooth) choice of a cotangent vector at each point. More formally, a vector field is a section of the tangent bundle TM and so, as we'll see, a 1-form is a section of the cotangent bundle T^*M .

Example 2.1.9. Suppose $M = \mathbb{R}^n$ endowed with its usual dot product. This induces an inner product on each tangent space $T_p \mathbb{R}^n$, which we will (confusingly, but in keeping with the usual practice in Riemannian geometry) denote by g . Specifically, we define g_p to be an inner product on $T_p \mathbb{R}^n$ given by $g_p(u, v) = u \cdot v$ for any $u, v \in T_p \mathbb{R}^n$.

Now, let $X \in \mathfrak{X}(\mathbb{R}^n)$. Then X corresponds to a unique $\alpha \in \Omega^1(\mathbb{R}^n)$ defined as follows: for $Y \in \mathfrak{X}(\mathbb{R}^n)$ and $p \in \mathbb{R}^n$, $\alpha(Y) \in C^\infty(\mathbb{R}^n)$ is given by

$$(\alpha(Y))(p) := g_p(X(p), Y(p))$$

(again, this is just the dot product of the vector $X(p)$ with the vector $Y(p)$). Then α is certainly a map $\mathfrak{X}(\mathbb{R}^n) \rightarrow C^\infty(\mathbb{R}^n)$, and it's smooth because X and g_p are. It's trivially alternating, and it's linear since g_p is. So $\alpha \in \Omega^1(\mathbb{R}^n)$.

In fact, any $\beta \in \Omega^1(\mathbb{R}^n)$ can be written in this way: at each $p \in \mathbb{R}^n$, β determines a linear functional $\beta_p \in (T_p \mathbb{R}^n)^*$. But then the Riesz Representation Theorem tells us that there exists $u_p \in T_p \mathbb{R}^n$ so that $\beta_p(v) = g_p(u_p, v)$ for all $v \in T_p \mathbb{R}^n$. In turn, we can define a vector field $U \in \mathfrak{X}(\mathbb{R}^n)$ by $U(p) := u_p$, and the smoothness of β will turn out to imply the smoothness of U .

This all tells us that g determines an isomorphism $\flat: \mathfrak{X}(\mathbb{R}^n) \rightarrow \Omega^1(\mathbb{R}^n)$ given by

$$X^\flat(Y) = g(X, Y).$$

In fact, there's nothing special about \mathbb{R}^n in the above. The same holds on any manifold M when g is a choice of *Riemannian metric* on M .³ Since it will turn out that we can always put a Riemannian metric on any manifold, this will tell us that, for any M ,

$$\Omega^1(M) \cong \mathfrak{X}(M)$$

either as vector spaces or as $C^\infty(M)$ -modules. You should view this isomorphism as the differential geometry analog of the isomorphism between a Hilbert space and its dual.

Example 2.1.10. We'll define $\omega \in \Omega^2(\mathbb{R}^2)$ as follows. Since we have global coordinates on \mathbb{R}^2 , we can write any $v \in T_p \mathbb{R}^2$ as

$$v = a \frac{\partial}{\partial x} + b \frac{\partial}{\partial y}$$

where $\left\{ \frac{\partial}{\partial x}, \frac{\partial}{\partial y} \right\}$ is just the standard basis written in the style of our local coordinate bases,⁴ and so any vector field

$$X(p) = a(p) \frac{\partial}{\partial x} + b(p) \frac{\partial}{\partial y}$$

where now $a, b \in C^\infty(M)$.

If $X, Y \in \mathfrak{X}(\mathbb{R}^2)$ are given in coordinates by $X = a \frac{\partial}{\partial x} + b \frac{\partial}{\partial y}$ and $Y = c \frac{\partial}{\partial x} + d \frac{\partial}{\partial y}$, define

$$\omega(X, Y) = \omega\left(a \frac{\partial}{\partial x} + b \frac{\partial}{\partial y}, c \frac{\partial}{\partial x} + d \frac{\partial}{\partial y}\right) := ad - bc = \begin{vmatrix} a & b \\ c & d \end{vmatrix}.$$

Since I've written this as a determinant, it is obviously alternating and multilinear, though that can also be checked directly.

³More precise definition to come, but basically a smooth choice of inner product on each tangent space.

⁴Formally, $\frac{\partial}{\partial x}$ and $\frac{\partial}{\partial y}$ really depend on the base point p , but since they're all parallel, I'm omitting this dependence.

So this ω is really a 2-form on \mathbb{R}^2 , which we're going to write as $\omega = dx \wedge dy$. The way to read this notation is as follows: dx pairs with $\frac{\partial}{\partial x}$ to produce 1, and dy pairs with $\frac{\partial}{\partial y}$ to produce 1; in other words, $\{dx, dy\}$ is the dual basis to $\left\{\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right\}$. Moreover, the wedge symbol forces this to be alternating and multilinear, so that, for example

$$dx \wedge dy \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y} \right) = 1 \quad \text{but} \quad dx \wedge dy \left(\frac{\partial}{\partial y}, \frac{\partial}{\partial x} \right) = -1.$$

Notice that $dx \wedge dy$ simply returns the signed area of the quadrilateral spanned by whatever pair of vectors is fed into it.

Example 2.1.11. There was nothing special about dimension 2 in the above example. [Theorem 2.1.3](#) implies that determinants are essentially the only way to get n -forms on n -dimensional manifolds, which recall are supposed to be alternating, multilinear maps $\omega: \mathfrak{X}(M)^n \rightarrow C^\infty(M)$: at each $p \in M$, such an ω is some scalar multiple of “the determinant” (whatever that precisely means in this general setting). This scalar is allowed to vary as we move around, so it turns out that

$$\omega = f \, dVol_M,$$

where $dVol_M$ is the name we give to the n -form which is just the determinant (again, whatever that really means, and with the caveat that $dVol_M$ does not always exist) on each $T_p M$, and f is some smooth function.

While this is obviously far from a rigorous proof, this hopefully gives you some intuition to the fact that $\Omega^n(M) \cong C^\infty(M)$.⁵

Just as with the previous example, geometrically $dVol_M$ is returning the signed n -dimensional volume of the parallelepiped spanned by any n vectors fed into it. This explains the notation and the terminology: we call this form a *volume form*.

Example 2.1.12. Let \det be the determinant on \mathbb{R}^n , which I think of as an alternating, multilinear map $\det: (\mathbb{R}^n)^n \rightarrow \mathbb{R}$. In fact, I'll use the same notation to indicate the induced map $\mathfrak{X}(\mathbb{R}^n)^n \rightarrow C^\infty(\mathbb{R}^n)$ by applying the determinant at each point. As in the previous example, I can think of $\det \in \Omega^n(\mathbb{R}^n)$.

But now we're going to combine \det with a vector field to define an $(n-1)$ -form η . Specifically, suppose $X \in \mathfrak{X}(\mathbb{R}^n)$. We'll define η pointwise, so let $p \in \mathbb{R}^n$. Since η is going to be an $(n-1)$ -form, it is supposed to input $n-1$ tangent vectors at p and output a number. To do so, we'll just tack on $X(p)$ and plug into \det ; that is, for $v_1, \dots, v_{n-1} \in T_p \mathbb{R}^n$, define

$$\eta_p(v_1, \dots, v_{n-1}) := \det(X(p), v_1, \dots, v_{n-1}).$$

This is alternating and multilinear, and the smoothness of X will imply it is smooth, so $\eta \in \Omega^{n-1}(M)$. Moreover, every $(n-1)$ -form can be written in this way, so this implies that $\mathfrak{X}(\mathbb{R}^n)$ is isomorphic to $\Omega^{n-1}(\mathbb{R}^n)$.

Again, as in [Example 2.1.9](#), this turns out to work on general manifolds when we have an analog of \det , which is our volume form from [Example 2.1.11](#). In general, then, this will imply that, for an orientable n -manifold M , $\Omega^1(M) \cong \Omega^{n-1}(M)$.

⁵Strictly speaking, this argument only works when M is orientable, as we will see.

2.2 Differential Forms and Vector Calculus

Suppose we restrict to the case of a (Riemannian) 3-manifold; that is, a 3-dimensional manifold M endowed with a Riemannian metric g (again, a Riemannian metric is a [smooth] choice of inner product g_p on each tangent space; if this makes you uncomfortable, just assume $M = \mathbb{R}^3$ and we just have the standard dot product on each tangent space). This section is a bit weird and definitely very informal: I'm using a bunch of stuff that I haven't actually defined yet. But the point is to try to say that concepts like *differential forms* and *exterior derivatives* that we will eventually define carefully are just generalizations of things you already understand very well.

From Examples 2.1.5–2.1.12, we know that

$$\Omega^0(M) = C^\infty(M) \cong \Omega^3(M), \quad \Omega^1(M) \cong \mathfrak{X}(M) \cong \Omega^2(M), \quad \text{and} \quad \Omega^k(M) = \{0\} \text{ for } k > 3.$$

So anything we can say about differential forms on M must be expressible just in terms of vector fields and functions. We know from vector calculus that there are various differentiation operators on functions and vector fields, namely div , grad , and curl , so what do these mean at the level of forms.

Or, turning it around, are there operations on forms which fill in the squares in the following diagram and makes them commute?

$$\begin{array}{ccccccc} \Omega^0(M) & & \Omega^1(M) & & \Omega^2(M) & & \Omega^3(M) \\ \parallel & & \cong \uparrow b & & \cong \uparrow & & \cong \uparrow \\ C^\infty(M) & \xrightarrow{\nabla} & \mathfrak{X}(M) & \xrightarrow{\nabla \times} & \mathfrak{X}(M) & \xrightarrow{\nabla \cdot} & C^\infty(M) \end{array}$$

First, we (hopefully) recall from vector calculus that the gradient of a function is a vector field, which can be computed in local coordinates as

$$\nabla f = \sum_{i=1}^3 \frac{\partial f}{\partial x_i} \frac{\partial}{\partial x_i}.$$

Exercise 2.2.1. If it's not clear, make the effort to connect this to the gradient you're familiar with.

Then the corresponding 1-form $(\nabla f)^b$ is defined at $p \in M$ by

$$(\nabla f)_p^b(Y) = g_p(\nabla f, Y) = g_p \left(\sum_{i=1}^3 \frac{\partial f}{\partial x_i} \frac{\partial}{\partial x_i}, \sum_{i=1}^3 b_i(p) \frac{\partial}{\partial x_i} \right) = \sum_{i,j=1}^3 f_{x_i}(p) b_j(p) g_p \left(\frac{\partial}{\partial x_i}, \frac{\partial}{\partial x_j} \right),$$

where $Y = \sum_{i=1}^3 b_i \frac{\partial}{\partial x_i}$ in local coordinates and I've written f_{x_i} as a shorthand for $\frac{\partial f}{\partial x_i}$.

If we were able to arrange it so that the local coordinate basis $\left\{ \frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \frac{\partial}{\partial x_3} \right\}$ were orthonormal with respect to g_p , then this would simplify as

$$(\nabla f)_p^b(Y) = \sum_{i=1}^3 f_{x_i}(p) b_i(p) = \sum_{i=1}^3 b_i(p) \frac{\partial f}{\partial x_i}(p) = \left(\left(\sum_{i=1}^3 b_i \frac{\partial}{\partial x_i} \right) f \right)(p) = (Yf)(p) = df_p(Y)$$

by Lemma 1.3.4. In other words, $(\nabla f)^b = df$.

Recall (Example 2.1.8) that at each point a 1-form corresponds to a cotangent vector, so the local coordinate basis $\left\{\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \frac{\partial}{\partial x_3}\right\}$ induces a dual basis which we call $\{dx_1, dx_2, dx_3\}$ for $(T_p M)^*$ defined by

$$dx_i \left(\frac{\partial}{\partial x_j} \right) = \delta_{ij}.$$

In these terms,

$$df = \frac{\partial f}{\partial x_1} dx_1 + \frac{\partial f}{\partial x_2} dx_2 + \frac{\partial f}{\partial x_3} dx_3.$$

In other words, in the presence of a Riemannian metric,⁶ computing the differential of a function on a 3-manifold (in fact, this works equally well on an n -manifold) corresponds to taking the gradient of the function.

Turning to curl, recall that, if $X \in \mathfrak{X}(M)$ is written in local coordinates as $X = a_1 \frac{\partial}{\partial x_1} + a_2 \frac{\partial}{\partial x_2} + a_3 \frac{\partial}{\partial x_3}$, then

$$\nabla \times X = \left(\frac{\partial a_3}{\partial x_2} - \frac{\partial a_2}{\partial x_3} \right) \frac{\partial}{\partial x_1} + \left(\frac{\partial a_1}{\partial x_3} - \frac{\partial a_3}{\partial x_1} \right) \frac{\partial}{\partial x_2} + \left(\frac{\partial a_2}{\partial x_1} - \frac{\partial a_1}{\partial x_2} \right) \frac{\partial}{\partial x_3}.$$

As discussed in Example 2.1.12, this corresponds to an element of $\omega^{3-1}(M) = \Omega^2(M)$ given by

$$\left(\frac{\partial a_3}{\partial x_2} - \frac{\partial a_2}{\partial x_3} \right) dx_2 \wedge dx_3 + \left(\frac{\partial a_1}{\partial x_3} - \frac{\partial a_3}{\partial x_1} \right) dx_3 \wedge dx_1 + \left(\frac{\partial a_2}{\partial x_1} - \frac{\partial a_1}{\partial x_2} \right) dx_1 \wedge dx_2.$$

For reasons to be explained later, I'm going to call this 2-form $\star(\nabla \times X)^\flat$.

So now what is the operation on $X^\flat = a_1 dx_1 + a_2 dx_2 + a_3 dx_3$ which would have produced $\star(\nabla \times X)^\flat$? The idea is to compute the differential (which is a 1-form) of each coefficient function, and combine it with the corresponding dx_i in a particular way:

$$\begin{aligned} d(X^\flat) &= d(a_1 dx_1 + a_2 dx_2 + a_3 dx_3) \\ &= (da_1) \wedge dx_1 + (da_2) \wedge dx_2 + (da_3) \wedge dx_3 \\ &= \left(\frac{\partial a_1}{\partial x_1} dx_1 + \frac{\partial a_1}{\partial x_2} dx_2 + \frac{\partial a_1}{\partial x_3} dx_3 \right) \wedge dx_1 + \left(\frac{\partial a_2}{\partial x_1} dx_1 + \frac{\partial a_2}{\partial x_2} dx_2 + \frac{\partial a_2}{\partial x_3} dx_3 \right) \wedge dx_2 \\ &\quad + \left(\frac{\partial a_3}{\partial x_1} dx_1 + \frac{\partial a_3}{\partial x_2} dx_2 + \frac{\partial a_3}{\partial x_3} dx_3 \right) \wedge dx_3 \\ &= 0 + \frac{\partial a_1}{\partial x_2} dx_2 \wedge dx_1 + \frac{\partial a_1}{\partial x_3} dx_3 \wedge dx_1 + \frac{\partial a_2}{\partial x_1} dx_1 \wedge dx_2 + 0 + \frac{\partial a_2}{\partial x_3} dx_3 \wedge dx_2 \\ &\quad + \frac{\partial a_3}{\partial x_1} dx_1 \wedge dx_3 + \frac{\partial a_3}{\partial x_2} dx_2 \wedge dx_3 + 0 \\ &= \left(\frac{\partial a_3}{\partial x_2} - \frac{\partial a_2}{\partial x_3} \right) dx_2 \wedge dx_3 + \left(\frac{\partial a_1}{\partial x_3} - \frac{\partial a_3}{\partial x_1} \right) dx_3 \wedge dx_1 + \left(\frac{\partial a_2}{\partial x_1} - \frac{\partial a_1}{\partial x_2} \right) dx_1 \wedge dx_2 \end{aligned}$$

where the rule is that $dx_i \wedge dx_j = -dx_j \wedge dx_i$, which is just the alternating condition on forms. So, if you buy the above manipulations on some level, there is a fairly straightforward generalization of the differential which corresponds exactly to the curl in this setting.

⁶which turns out to be necessary to define the gradient anyway

Finally, if $X \in \mathfrak{X}(M)$ is written in local coordinates as $X = a_1 \frac{\partial}{\partial x_1} + a_2 \frac{\partial}{\partial x_2} + a_3 \frac{\partial}{\partial x_3}$, then the divergence is

$$\nabla \cdot X = \frac{\partial a_1}{\partial x_1} + \frac{\partial a_2}{\partial x_2} + \frac{\partial a_3}{\partial x_3},$$

which is a function. But the isomorphism $C^\infty(M) \cong \Omega^3(M)$ described in [Example 2.1.11](#) says that this function corresponds to some 3-form

$$\left(\frac{\partial a_1}{\partial x_1} + \frac{\partial a_2}{\partial x_2} + \frac{\partial a_3}{\partial x_3} \right) dVol_M = \left(\frac{\partial a_1}{\partial x_1} + \frac{\partial a_2}{\partial x_2} + \frac{\partial a_3}{\partial x_3} \right) dx_1 \wedge dx_2 \wedge dx_3$$

because for an appropriate choice of coordinates I can write $dVol_M = dx_1 \wedge dx_2 \wedge dx_3$.

Can we get to this 3-form by applying our generalized differential to the 2-form $\star X^\flat = a_1 dx_2 \wedge dx_3 + a_2 dx_3 \wedge dx_1 + a_3 dx_1 \wedge dx_2$? Yes!

$$\begin{aligned} d(\star X^\flat) &= d(a_1 dx_2 \wedge dx_3 + a_2 dx_3 \wedge dx_1 + a_3 dx_1 \wedge dx_2) \\ &= (da_1) \wedge dx_2 \wedge dx_3 + (da_2) \wedge dx_3 \wedge dx_1 + (da_3) \wedge dx_1 \wedge dx_2 \\ &= \left(\frac{\partial a_1}{\partial x_1} dx_1 + \frac{\partial a_1}{\partial x_2} dx_2 + \frac{\partial a_1}{\partial x_3} dx_3 \right) \wedge dx_2 \wedge dx_3 + \left(\frac{\partial a_2}{\partial x_1} dx_1 + \frac{\partial a_2}{\partial x_2} dx_2 + \frac{\partial a_2}{\partial x_3} dx_3 \right) \wedge dx_3 \wedge dx_1 \\ &\quad + \left(\frac{\partial a_3}{\partial x_1} dx_1 + \frac{\partial a_3}{\partial x_2} dx_2 + \frac{\partial a_3}{\partial x_3} dx_3 \right) \wedge dx_1 \wedge dx_2 \\ &= \frac{\partial a_1}{\partial x_1} dx_1 \wedge dx_2 \wedge dx_3 + 0 + 0 + 0 + \frac{\partial a_2}{\partial x_2} dx_2 \wedge dx_3 \wedge dx_1 + 0 + 0 + \frac{\partial a_3}{\partial x_3} dx_3 \wedge dx_1 \wedge dx_2 \\ &= \frac{\partial a_1}{\partial x_1} dx_1 \wedge dx_2 \wedge dx_3 + \frac{\partial a_2}{\partial x_2} dx_1 \wedge dx_2 \wedge dx_3 + \frac{\partial a_3}{\partial x_3} dx_1 \wedge dx_2 \wedge dx_3 \\ &= \left(\frac{\partial a_1}{\partial x_1} + \frac{\partial a_2}{\partial x_2} + \frac{\partial a_3}{\partial x_3} \right) dx_1 \wedge dx_2 \wedge dx_3 \end{aligned}$$

The upshot is that we've now filled in the diagram:

$$\begin{array}{ccccccc} \Omega^0(M) & \xrightarrow{d} & \Omega^1(M) & \xrightarrow{d} & \Omega^2(M) & \xrightarrow{d} & \Omega^3(M) \\ \parallel & & \cong \uparrow \flat & & \cong \uparrow \star^\flat & & \cong \uparrow \cdot dVol_M \\ C^\infty(M) & \xrightarrow{\nabla} & \mathfrak{X}(M) & \xrightarrow{\nabla \times} & \mathfrak{X}(M) & \xrightarrow{\nabla \cdot} & C^\infty(M) \end{array}$$

Of course, I haven't really given a rigorous definition of d (which is called the *exterior derivative*), nor of \wedge (the *wedge product*), but, as we will see, there is a natural, coordinate-free way of defining these things for arbitrary differential forms on arbitrary manifolds which specializes to the calculations above in local coordinates (and hence corresponds to gradient and divergence on manifolds of arbitrary dimension).

Hopefully the fact that div, grad, and curl are all essentially the same operator in this framework helps convince you that differential forms are useful and important. In fact, it gets even better: in this language, the Fundamental Theorem of Calculus, Green's Theorem, and the Divergence Theorem all turn out to be special cases of the same theorem: *Stokes' Theorem for differential forms*.

More than just generalizing essentially all of vector calculus, differential forms also encode topological information (in the form of cohomology). To see this, look again at the diagram above:

- (i) We know from vector calculus that $\nabla \times (\nabla f) = 0$ for any $f \in C^\infty(M)$ and $\nabla \cdot (\nabla \times X) = 0$ for any $X \in \mathfrak{X}(M)$. Since the diagram commutes, this implies that $d \circ d = 0$ regardless of whether you start in $\Omega^0(M)$ or $\Omega^1(M)$ (and in fact this is still true if you start in $\Omega^2(M)$ or $\Omega^3(M)$ since $\Omega^4(M) = \{0\}$). In other words,

$$\Omega^0(M) \xrightarrow{d} \Omega^1(M) \xrightarrow{d} \Omega^2(M) \xrightarrow{d} \Omega^3(M)$$

is a (co)chain complex. More generally, if we add indices to the exterior derivatives to indicate where they start and end,

$$\Omega^0(M) \xrightarrow{d_0} \Omega^1(M) \xrightarrow{d_1} \dots \xrightarrow{d_{n-2}} \Omega^{n-1}(M) \xrightarrow{d_{n-1}} \Omega^n(M)$$

is always a cochain complex on any n -manifold (i.e., $d_k \circ d_{k-1} = 0$ for all k). The usual thing you do when you have a (co)chain complex is to take the (co)homology, and that's also useful in this case. Doing so yields the *de Rham cohomology groups*

$$H_{\text{dR}}^k(M) := \frac{\ker d_k}{\text{im } d_{k-1}}.$$

In fact, as usual with cohomology, the cohomology groups fit together into a graded ring, and the product operation corresponds to the wedge product \wedge . And de Rham cohomology will turn out to be isomorphic to singular (or simplicial) cohomology with real coefficients.

- (ii) In our 3-manifold example, we saw that $\Omega^0(M) = C^\infty(M) \cong \Omega^3(M)$ and $\Omega^1(M) \cong \mathfrak{X}(M) \cong \Omega^2(M)$. In general, if M is an n -dimensional manifold, then it will turn out that $\Omega^k(M) \cong \Omega^{n-k}(M)$ for all $k = 0, \dots, n$. Even better, this isomorphism descends to the de Rham cohomology groups and we will have that

$$H_{\text{dR}}^k(M) \cong H_{\text{dR}}^{n-k}(M)$$

for all k , which is the appropriate version of *Poincaré duality* for de Rham cohomology.

2.3 Tensor Algebras and Tensor Fields

In order to give a more rigorous definition of differential forms (and in particular to get them to form a graded algebra), as well as more general tensor fields, we need to do some (multilinear) algebra and talk about tensor algebras and exterior algebras.

Before diving in, let me just give you my perspective on tensors. Basically, the point is that tensors are the right tool for turning multilinear algebra into linear algebra. More precisely, suppose we have three vector spaces U , V , and W , and a map $F : U \times V \rightarrow W$ which is multilinear (or, really, bilinear in this case). Again, this just means that F is linear in each factor:

$$F(au_1 + bu_2, v) = aF(u_1, v) + bF(u_2, v) \quad \text{and} \quad F(u, cv_1 + dv_2) = cF(u, v_1) + dF(u, v_2).$$

Example 2.3.1. Any choice of inner product on \mathbb{R}^n defines a bilinear map $\mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ by $(u, v) \mapsto \langle u, v \rangle$. Here $U = V = \mathbb{R}^n$ and $W = \mathbb{R}$.

Example 2.3.2. Let $U = \mathbb{R}^n$, $V = (\mathbb{R}^n)^{n-1}$ and define a map $\mathbb{R}^n \times (\mathbb{R}^n)^{n-1} \cong (\mathbb{R}^n)^n \rightarrow \mathbb{R}$ by

$$(u, (v_1, \dots, v_{n-1})) \mapsto \det [u \quad v_1 \quad \dots \quad v_{n-1}].$$

This is also bilinear.

Example 2.3.3. Let $U = V = W = \mathbb{R}^3$ and define the map $\mathbb{R}^3 \times \mathbb{R}^3 \rightarrow \mathbb{R}^3$ by $(u, v) \mapsto u \times v$, which is again bilinear.

Returning to the general setting, we have vector space U , V , and W and a multilinear map $F : U \times V \rightarrow W$. Now suppose $w_0 \in W$ and we want to solve a problem of the form $F(u, v) = w_0$. Since F is multilinear, you might hope that this is somehow just a linear algebra problem, which would be solvable, at least in principle. But, at least as stated, this is very much *not* a linear problem...

Example 2.3.4. Continuing with Example 2.3.3, where $F : \mathbb{R}^3 \times \mathbb{R}^3 \rightarrow \mathbb{R}^3$ is given by $F(u, v) = u \times v$, suppose we want to solve $F(u, v) = (x_0, y_0, z_0)$. Since $\mathbb{R}^3 \times \mathbb{R}^3 \cong \mathbb{R}^6$, we can think of F as a map $\mathbb{R}^6 \rightarrow \mathbb{R}^3$ given by

$$(u_1, u_2, u_3, v_1, v_2, v_3) \mapsto u \times v = (u_2 v_3 - u_3 v_2, u_3 v_1 - u_1 v_3, u_1 v_2 - u_2 v_1).$$

So our problem is to solve

$$(u_2 v_3 - u_3 v_2, u_3 v_1 - u_1 v_3, u_1 v_2 - u_2 v_1) = (x_0, y_0, z_0)$$

for $(u_1, u_2, u_3, v_1, v_2, v_3)$, which is obviously a *quadratic* system of equations, not a linear system.

To reiterate, for me⁷ the point of tensor products is to turn multilinear maps and problems into linear maps and problems.

Definition 2.3.5. Given vector spaces U and V , define the *tensor product* $U \otimes V$ to be the vector space whose elements are linear combinations of terms of the form $u \otimes v$ for $u \in U$ and $v \in V$ so that

- (i) $(u_1 + u_2) \otimes v = u_1 \otimes v + u_2 \otimes v$
- (ii) $u \otimes (v_1 + v_2) = u \otimes v_1 + u \otimes v_2$
- (iii) $a(u \otimes v) = (au) \otimes v = u \otimes (av)$

for any $u, u_1, u_2 \in U$, $v, v_1, v_2 \in V$ and $a \in \mathbb{R}$.

Example 2.3.6. If $U = \mathbb{R}^m$ and $V = \mathbb{R}^n$, then we can represent $u \otimes v$ by the $m \times n$ matrix

$$uv^T = \begin{bmatrix} u_1 \\ \vdots \\ u_m \end{bmatrix} \begin{bmatrix} v_1 & \dots & v_n \end{bmatrix} = [u_i v_j]_{i,j}.$$

Here's the key theorem (or, if you start from a different perspective, the below theorem is the definition):

Theorem 2.3.7 (Universal Property of the Tensor Product). *If $\phi : U \times V \rightarrow U \otimes V$ is the map given by $(u, v) \mapsto u \otimes v$ and $F : U \times V \rightarrow W$ is bilinear, then there exists a unique linear map $\tilde{F} : U \otimes V \rightarrow W$ making the following diagram commute:*

⁷I want to emphasize here that this is just my perspective: I'm not claiming that everybody would agree with this characterization.

$$\begin{array}{ccc} U \times V & \xrightarrow{F} & W \\ \phi \downarrow & \nearrow \tilde{F} & \\ U \otimes V & & \end{array}$$

Moreover, this uniquely characterizes $U \otimes V$: any vector space satisfying this property must be isomorphic to $U \otimes V$.

Example 2.3.8. Continuing Examples 2.3.3 and 2.3.4, recall that $F: \mathbb{R}^3 \times \mathbb{R}^3 \rightarrow \mathbb{R}^3$ is given by $F(u, v) = u \times v$, and Theorem 2.3.7 tells us there must be a linear map $\tilde{F}: \mathbb{R}^3 \otimes \mathbb{R}^3 \rightarrow \mathbb{R}^3$ so that

$$(\tilde{F} \circ \phi)(u, v) = F(u, v) = u \times v.$$

Representing $u \otimes v$ by uv^T as in Example 2.3.6, we see that $u \otimes v$ corresponds to the 3×3 matrix

$$\begin{bmatrix} u_1v_1 & u_1v_2 & u_1v_3 \\ u_2v_1 & u_2v_2 & u_2v_3 \\ u_3v_1 & u_3v_2 & u_3v_3 \end{bmatrix}.$$

In general, $\mathbb{R}^3 \otimes \mathbb{R}^3 \cong \text{Mat}_{3 \times 3}(\mathbb{R}) \cong \mathbb{R}^9$, so we can flatten this matrix to get a representation by a 9-dimensional vector, namely

$$\begin{bmatrix} u_1v_1 \\ u_1v_2 \\ u_1v_3 \\ u_2v_1 \\ u_2v_2 \\ u_2v_3 \\ u_3v_1 \\ u_3v_2 \\ u_3v_3 \end{bmatrix}.$$

So then $\tilde{F}: \mathbb{R}^9 \rightarrow \mathbb{R}^3$ is supposed to be linear, and hence must be represented by a 3×9 matrix A so that

$$A \begin{bmatrix} u_1v_1 \\ u_1v_2 \\ u_1v_3 \\ u_2v_1 \\ u_2v_2 \\ u_2v_3 \\ u_3v_1 \\ u_3v_2 \\ u_3v_3 \end{bmatrix} = \begin{bmatrix} u_2v_3 - u_3v_2 \\ u_3v_1 - u_1v_3 \\ u_1v_2 - u_2v_1 \end{bmatrix}.$$

Written out in this excruciating detail, it's now pretty obvious what A has to be:

$$A = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Proposition 2.3.9. If u_1, \dots, u_m is a basis for U and v_1, \dots, v_n is a basis for V , then $\{e_i \otimes f_j\}_{i,j}$ is a basis for $U \otimes V$. In particular, $\dim U \otimes V = mn$.

Clearly, we've paid a price for turning bilinear maps into linear maps: whereas $\dim(U \times V) = m + n$, we have $\dim(U \otimes V) = mn$, which is usually a lot bigger. So we turn multilinear algebra into linear algebra at the cost of having to work in much higher dimensions.

Notation. We will use $V^{\otimes k}$ to denote the k -fold tensor product of a vector space V with itself; that is

$$\underbrace{V \otimes \cdots \otimes V}_k = V \otimes (V \otimes (\cdots \otimes (V \otimes V)))$$

Definition 2.3.10. The *tensor space of type (r, s)* associated with a vector space V is the vector space

$$V_r^s := V^{\otimes r} \otimes (V^*)^{\otimes s}.$$

Elements of V_r^s are called (r, s) -*tensors*.

Example 2.3.11. An inner product $\langle \cdot, \cdot \rangle$ on V is a symmetric $(0, 2)$ -tensor on V . Why? Well, $\langle \cdot, \cdot \rangle$ defines a bilinear map $V \times V \rightarrow \mathbb{R}$, and hence by [Theorem 2.3.7](#) a (unique) linear map $V \otimes V \rightarrow \mathbb{R}$. In other words, an inner product determines an element of $(V \otimes V)^* \cong V^* \otimes V^* = V_0^2$.

Exercise 2.3.12. Prove that $(V \otimes V)^* \cong V^* \otimes V^*$

Example 2.3.13. A linear transformation $F: V \rightarrow V$ (or, if you like, a square matrix) can be interpreted as a $(1, 1)$ -tensor. Why? F induces a bilinear map $V^* \times V \rightarrow \mathbb{R}$ as follows: for $(\rho, v) \in V^* \times V$,

$$(\rho, v) \mapsto \rho(F(v)).$$

Again, [Theorem 2.3.7](#) implies that this bilinear map corresponds to a linear map $V^* \otimes V \rightarrow \mathbb{R}$; that is, an element of the dual space $(V^* \otimes V)^* \cong V \otimes V^* = V_1^1$.

Definition 2.3.14. The direct sum $\mathcal{T}(V) := \bigoplus_{r,s \geq 0} V_r^s$, where V_0^0 is the ground field of V , is called the *tensor algebra* of V .

Remark 2.3.15. Terminology varies. Often *tensor algebra* only refers to the algebra of $(\cdot, 0)$ -tensors, namely $\bigoplus_{r \geq 0} V_r^0 = \bigoplus_{r \geq 0} V^{\otimes r}$.

In general, $\mathcal{T}(V)$ is noncommutative, associative, and (bi-)graded.

Definition 2.3.16. The *tensor bundle of type (r, s)* over a manifold M is

$$\mathcal{T}_r^s(M) := \bigsqcup_{p \in M} (T_p M)_r^s,$$

which has a projection map $\pi: \mathcal{T}_r^s(M) \rightarrow M$ sending a tensor based at a point to the point.

Exercise 2.3.17. If M is an n -dimensional manifold, show that $\mathcal{T}_r^s(M)$ is a smooth manifold of dimension $n + n^{r+s}$.

A *tensor field* of type (r, s) on M is a smooth section of the tensor bundle $\mathcal{T}_r^s(M)$; i.e., a smooth map $\tau: M \rightarrow \mathcal{T}_r^s(M)$ so that $\tau(p) \in (T_p M)^s_r$.

Example 2.3.18. A vector field on M is (by definition) a $(1, 0)$ -tensor field on M .

Example 2.3.19. A 1-form on M is (again, by definition) a $(0, 1)$ -tensor field.

Example 2.3.20. A Riemannian metric (which we've informally defined to be a smooth choice of inner product on each tangent space) is a (symmetric) $(0, 2)$ -tensor field on M . More precisely:

Definition 2.3.21. A *Riemannian metric* on a manifold M is a smooth $(0, 2)$ -tensor field g on M so that, for each $p \in M$, $g_p = g(p)$ satisfies the axioms of an inner product on $T_p(M)$; that is, in addition to being bilinear (which is guaranteed by the fact it is a $(0, 2)$ -tensor), it must be symmetric ($g_p(u, v) = g_p(v, u)$) and positive-definite ($g_p(v, v) \geq 0$ with equality if and only if $v = 0$).

Note: when I write something like $g_p(u, v)$ I'm technically thinking of g_p as the associated bilinear map $T_p M \times T_p M \rightarrow \mathbb{R}$, but in coordinates we also often use notation like $g = \sum_{i,j} g_{ij} dx_i \otimes dx_j$, which more closely matches up with the idea of g being a $(0, 2)$ -tensor field.

Example 2.3.22. Consider the Riemannian metric $g = \frac{1}{y^2}(dx \otimes dx + dy \otimes dy)$ (also commonly written as $ds^2 = \frac{1}{y^2}(dx^2 + dy^2)$) on $H = \{(x, y) \in \mathbb{R}^2 : y > 0\}$.⁸ This is the Poincaré half-plane model of the hyperbolic plane. In general, we can compute the length of a curve $\gamma(t) = (\gamma_1(t), \gamma_2(t))$ by integrating the length of $\gamma'(t)$ along the curve. Since $\gamma'(t) \in T_{\gamma(t)} H$, its length is

$$\sqrt{g_{\gamma(t)}(\gamma'(t), \gamma'(t))} = \sqrt{\frac{\gamma'(t) \cdot \gamma'(t)}{\gamma_2(t)^2}} = \frac{\|\gamma'(t)\|_E}{\gamma_2(t)},$$

where \cdot is the standard dot product on the plane and $\|\cdot\|_E$ is the Euclidean norm (I can just use $\gamma_2(t)$ in the denominator, rather than $|\gamma_2(t)|$, because the second coordinate is always positive in H).

It will turn out that the geodesics in this metric are the vertical lines and semicircles centered at points on the x -axis; see [Figure 2.1](#). Let's compute distances between points on these geodesics.

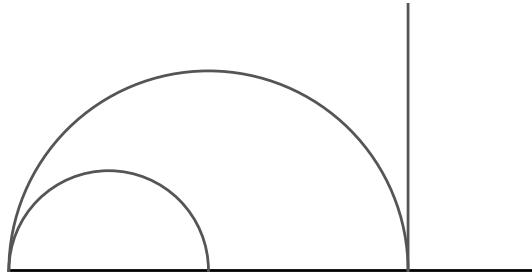


Figure 2.1: Some geodesics in the Poincaré upper half-plane model of the hyperbolic plane.

⁸In undergraduate differential geometry courses like MATH 474, the Riemannian metric is more often called the First Fundamental Form.

For the vertical lines, we can parametrize the line through $(x_0, 0)$ by $\gamma(t) = (x_0, t)$. So then $\gamma'(t) = (0, 1)$ and the distance between (x_0, y_0) and (x_0, y_1) is

$$\int_{y_0}^{y_1} \sqrt{g_{\gamma(t)}(\gamma'(t), \gamma'(t))} dt = \int_{y_0}^{y_1} \frac{1}{t} dt = \ln y_1 - \ln y_0 = \ln \left(\frac{y_1}{y_0} \right).$$

Notice, in particular, that the x -axis is infinitely far away from any point in H .

On the other hand, the semicircle of radius r centered at $(x_0, 0)$ can be parametrized by $\gamma(t) = (x_0 + r \cos t, r \sin t)$. Now $\gamma'(t) = r(-\sin t, \cos t)$, so the distance between $\gamma(\theta_0)$ and $\gamma(\theta_1)$ is

$$\int_{\theta_0}^{\theta_1} \sqrt{g_{\gamma(t)}(\gamma'(t), \gamma'(t))} dt = \int_{\theta_0}^{\theta_1} \frac{r}{r \sin t} dt = \int_{\theta_0}^{\theta_1} \csc t dt = \operatorname{arctanh}(\cos \theta_0) - \operatorname{arctanh}(\cos \theta_1).$$

2.4 Exterior Algebras and Differential Forms

We just saw that 1-forms are $(0, 1)$ -tensor fields, but in general it's not quite natural to think of k -forms as tensor fields because we need to incorporate the alternating condition. We could do this by considering the subset of alternating tensor fields, but it turns out to be more natural to take a quotient of the tensor algebra to get something called the *exterior algebra*.

Specifically, just as the tensor product is a way of turning multilinear maps into linear maps, the *exterior product* will turn out to be the way to turn alternating multilinear maps into linear maps. Even though we haven't actually defined the exterior product yet, it might seem silly to introduce another new product; after all, alternating multilinear maps are multilinear so the tensor product *already* turns them into linear maps.

That being the case, the exterior product does this in much lower dimension, so it's still worthwhile. At the most extreme level, you can see this because the natural space in which the tensor product lives (as a binary operation) is the tensor algebra $\mathcal{T}(V)$ (or, if you like, only the $(r, 0)$ part $V_0^0 \otimes V_1^0 \otimes V_2^0 \otimes \dots$), which is always infinite-dimensional. Whereas the exterior product will be a binary operation on the exterior algebra, which will turn out to be finite-dimensional if V is. And, if you weren't already aware, finite-dimensional vector spaces are *much, much nicer* than infinite-dimensional vector spaces: it's always worthwhile to make the effort to work in finite dimensions if at all possible.

Definition 2.4.1. Let $C(V) = \bigoplus_{k \geq 0} V_k^0$ be the algebra of $(\cdot, 0)$ tensors on a vector space V and let $\mathcal{I}(V)$ be the two-sided ideal generated by elements of the form $v \otimes v$ for $v \in V$. Then the *exterior algebra* of V is the quotient

$$\wedge(V) := C(V)/\mathcal{I}(V).$$

This is a graded algebra with product denoted by \wedge , which is just the product induced by \otimes : the residue class of $v_1 \otimes v_2 \otimes \dots \otimes v_k$ is denoted $v_1 \wedge v_2 \wedge \dots \wedge v_k$.

The grading is given by the degree of the wedge product, since

$$\wedge(V) = \bigoplus_{k \geq 0} \wedge^k(V),$$

where $\wedge^k(V) = V_k^0/\mathcal{I}_k$, and $\mathcal{I}_k = \mathcal{I}(V) \cap V_k^0$.

Proposition 2.4.2. If $a \in \wedge^k(V)$ and $b \in \wedge^\ell(V)$, then $b \wedge a = (-1)^{k\ell} a \wedge b$.

Exercise 2.4.3. Prove Proposition 2.4.2.

Corollary 2.4.4. $v_{\sigma(1)} \wedge \cdots \wedge v_{\sigma(k)} = \text{sgn}(\sigma)v_1 \wedge \cdots \wedge v_k$ for any permutation σ of $\{1, \dots, k\}$.

In particular, this says that we can always rearrange any wedge product to have the terms in any order we like. For example, if the terms are indexed (e.g., if they come from some ordered basis), we can always make the indices appear in increasing order at the cost of a sign.

Proposition 2.4.5. If $\{e_1, \dots, e_n\}$ is a basis for V , then

$$\{e_{i_1} \wedge \cdots \wedge e_{i_k} : 1 \leq i_1 < i_2 < \cdots < i_k \leq n\}$$

is a basis for $\wedge^k(V)$. In particular,

$$\dim \wedge^k(V) = \begin{cases} \binom{n}{k} & \text{if } 0 \leq k \leq n \\ 0 & \text{else.} \end{cases}$$

Hence,

$$\dim \wedge(V) = \sum_{k \geq 0} \dim \wedge^k(V) = \sum_{k=0}^n \binom{n}{k} = 2^n.$$

Exercise 2.4.6. Prove Proposition 2.4.5.

This is exactly the finite-dimensionality described above. Even if we're just interested in alternating, bilinear maps $V \times V \rightarrow W$, $\dim \wedge^2(V) = \binom{n}{2}$ is less than half of $\dim V^{\otimes 2} = n^2$, so working with the exterior power rather than the tensor power means we're doing linear algebra in half the dimensions, which is almost always worthwhile computationally.

As you would expect, the exterior products satisfy a universal property analogous to the one satisfied by tensor products:

Theorem 2.4.7 (Universal Property of the Exterior Product). *If $\phi: V^k \rightarrow \wedge^k(V)$ is the map given by $(v_1, \dots, v_k) \mapsto v_1 \wedge \cdots \wedge v_k$ and $F: V^k \rightarrow W$ is an alternating multilinear map, then there exists a unique linear map $\tilde{F}: \wedge^k(V) \rightarrow W$ making the following diagram commute:*

$$\begin{array}{ccc} \underbrace{V \times \cdots \times V}_k & \xrightarrow{F} & W \\ \phi \downarrow & \nearrow \tilde{F} & \\ \wedge^k(V) & & \end{array}$$

Again, this uniquely characterizes $\wedge^k(V)$ up to isomorphism.

Example 2.4.8. Suppose $V = \mathbb{R}^n$, and suppose $F: \mathbb{R}^n \rightarrow \mathbb{R}$ is alternating and multilinear. We know that this factors through a linear map $\wedge^n(\mathbb{R}^n) \rightarrow \mathbb{R}$; since $\dim \wedge^n \mathbb{R}^n = \binom{n}{n} = 1$, we know that $\wedge^n \mathbb{R}^n \cong \mathbb{R}$, and so (once we've chosen a basis for $\wedge^n \mathbb{R}^n$, which witnesses the isomorphism $\wedge^n \mathbb{R}^n \cong \mathbb{R}$), \tilde{F} has to be of the form $x \mapsto ax$ for some scalar a . We can already see that this is going to imply that F is unique up to the choice of the scalar a .

Let's work out what $\phi: (\mathbb{R}^n)^n \rightarrow \wedge^n \mathbb{R}^n$ is. Let e_1, \dots, e_n be the standard basis for \mathbb{R}^n , so that each $v_i = \sum_{j_i=1}^n a_{ij_i} e_{j_i}$. Then repeatedly applying Corollary 2.4.4 and the alternating property of the wedge product yields

$$\begin{aligned}\phi(v_1, \dots, v_n) &= \phi\left(\sum_{j_1=1}^n a_{1j_1} e_{j_1}, \dots, \sum_{j_n=1}^n a_{nj_n} e_{j_n}\right) \\ &= \left(\sum_{j_1=1}^n a_{1j_1} e_{j_1}\right) \wedge \cdots \wedge \left(\sum_{j_n=1}^n a_{nj_n} e_{j_n}\right) \\ &= \sum_{j_1=1}^n \cdots \sum_{j_n=1}^n a_{1j_1} \cdots a_{nj_n} e_{j_1} \wedge \cdots \wedge e_{j_n} \\ &= \sum_{\sigma \in S_n} \left[\left(\prod_{i=1}^n a_{i\sigma(i)} \right) e_{\sigma(1)} \wedge \cdots \wedge e_{\sigma(n)} \right] \\ &= \sum_{\sigma \in S_n} \left[\operatorname{sgn}(\sigma) \left(\prod_{i=1}^n a_{i\sigma(i)} \right) e_1 \wedge \cdots \wedge e_n \right] \\ &= \left[\sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \left(\prod_{i=1}^n a_{i\sigma(i)} \right) \right] e_1 \wedge \cdots \wedge e_n \\ &= \det [v_1 \quad \cdots \quad v_n] e_1 \wedge \cdots \wedge e_n,\end{aligned}$$

where we just recognized the Leibniz formula for the determinant in the last equality.

Since $\{e_1 \wedge \cdots \wedge e_n\}$ is our preferred basis for $\wedge^n \mathbb{R}^n$, inducing the isomorphism $\mathbb{R} \xrightarrow{\cong} \wedge^n \mathbb{R}^n$ given by $1 \mapsto e_1 \wedge \cdots \wedge e_n$, we see that

$$F(v_1, \dots, v_n) = (\tilde{F} \circ \phi)(v_1, \dots, v_n) = \tilde{F}(\phi(v_1, \dots, v_n)) = \tilde{F}(\det [v_1 \quad \cdots \quad v_n] e_1 \wedge \cdots \wedge e_n) = a \det [v_1 \quad \cdots \quad v_n]$$

for some $a \in \mathbb{R}$, which (re)proves Theorem 2.1.3, which says that the determinant is, up to scale, the unique alternating multilinear map $(\mathbb{R}^n)^n \rightarrow \mathbb{R}$.

If we let $W = \mathbb{R}$ (or whatever the base field is) in Theorem 2.4.7, then we see that alternating multilinear functionals on V^k correspond precisely to linear functionals on $\wedge^k V$; that is, to elements of the dual space $(\wedge^k V)^*$. But of course this is essentially how we “defined” differential k -forms in Definition 2.1.4: at each point a k -form should exactly be an alternating multilinear functional on the product of k copies of the tangent space at that point.

Definition 2.4.9. The *exterior k -bundle* on a manifold M is

$$\wedge^k(M) := \bigsqcup_{p \in M} \wedge^k((T_p M)^*)$$

and the *exterior algebra bundle* over M is

$$\wedge(M) := \bigsqcup_{p \in M} \wedge((T_p M)^*) .$$

Both of these are smooth manifolds.

A *differential k-form* on M is a smooth section of $\bigwedge_n^k(M)$; that is, a smooth map $\omega: M \rightarrow \bigwedge_n^k(M)$ so that $\omega(p) \in \bigwedge^k((T_p M)^*)$. The space of k -forms on M is denoted $\Omega^k(M)$, and we let $\Omega^*(M) := \bigoplus_{k=0}^n \Omega^k(M)$, which is the algebra of differential forms on M .

If you've been reading very closely, you might have noticed that I played fast and loose with parentheses and stars above. The discussion in the paragraph before [Definition 2.4.9](#) suggests that, if ω is a k -form (that is, an alternating, multilinear map $\mathfrak{X}(M)^k \rightarrow C^\infty(M)$ according to our informal definition), then at a point $p \in M$ we should expect that $\omega(p) \in (\bigwedge^k(T_p M))^*$. But I've just defined $\omega(p)$ as an element of $\bigwedge^k((T_p M)^*)$! So what's going on?

First of all, $\bigwedge^k((T_p M)^*)$ is (at least to me) conceptually nicer than $(\bigwedge^k(T_p M))^*$: it's an exterior power itself, rather than being the dual of an exterior power, and so it obviously fits into an exterior algebra, namely $\bigwedge((T_p M)^*)$. So that's the reason for wanting to define exterior bundles in terms of $\bigwedge^k((T_p M)^*)$ rather than $(\bigwedge^k(T_p M))^*$.

Of course, the reason I can do this is because they're isomorphic:

Lemma 2.4.10. *Suppose V is a finite-dimensional vector space. Then*

$$\bigwedge^k(V^*) \cong (\bigwedge^k V)^*,$$

and therefore

$$\bigwedge(V^*) \cong (\bigwedge(V))^*.$$

The strategy here is to find a nondegenerate pairing between $\bigwedge^k(V^*)$ and $\bigwedge^k(V)$. Recall that a *nondegenerate pairing* of vector space V and W is a bilinear map

$$(\cdot, \cdot): V \times W \rightarrow \mathbb{R}$$

with the property that, for any nonzero $w \in W$, there exists $v \in V$ so that $(v, w) \neq 0$ (and similarly when you swap the roles of v and w). Notice, in particular, that a pairing corresponds uniquely to a linear map $V \otimes W \rightarrow \mathbb{R}$; that is, an element of $(V \otimes W)^*$.

If there is a nondegenerate pairing of V and W , then it induces injective maps $\phi: V \rightarrow W^*$ and $\psi: W \rightarrow V^*$ given by

$$\phi(v)(w) := (v, w) =: \psi(w)(v).$$

Example 2.4.11. An inner product $\langle \cdot, \cdot \rangle$ on V is a nondegenerate pairing of V with itself: after all, if $v \neq 0$, then $\langle v, v \rangle \neq 0$ by the positivity of the inner product. Hence, the inner product induces an injective map $V \rightarrow V^*$ which turns out to be an isomorphism if V is finite-dimensional (or more generally if we take V^* to be the continuous dual space rather than the algebraic dual space); this isomorphism is the content of the Riesz representation theorem.

In finite dimensions, we know that $V \cong V^*$ and $W \cong W^*$, so the composition $V \xhookrightarrow{\phi} W^* \cong W \xhookrightarrow{\psi} V^*$ gives an injective (and therefore also surjective) map $V \rightarrow V^*$, which implies that ϕ and ψ had to be bijective, and therefore isomorphisms. In particular, this implies that $V \cong W^*$ (and therefore also $W \cong V^*$ and indeed $V \cong W$).

So the strategy for proving [Lemma 2.4.10](#) is to find a nondegenerate pairing of $\wedge^k(V^*)$ and $\wedge^k(V)$, which will imply the result.⁹

Proof of Lemma 2.4.10. To find the pairing $(\cdot, \cdot) : \wedge^k(V^*) \times \wedge^k(V) \rightarrow \mathbb{R}$, suppose first that $\alpha \in \wedge^k(V^*)$ and $b \in \wedge^k(V)$ are *decomposable*, meaning that $\alpha = \alpha_1 \wedge \cdots \wedge \alpha_k$ for $\alpha_1, \dots, \alpha_k \in V^*$ and $b = b_1 \wedge \cdots \wedge b_k$ for $b_1, \dots, b_k \in V$. Then define

$$(\alpha, b) := \det \left([\alpha_i(b_j)]_{i,j} \right). \quad (2.2)$$

Since arbitrary elements of $\wedge^k(V^*)$ and $\wedge^k(V)$ are linear combinations of decomposable elements, we define $(\cdot, \cdot) : \wedge^k(V^*) \times \wedge^k(V) \rightarrow \mathbb{R}$ by extending the above linearly.

It is straightforward (but somewhat tedious) to check that (\cdot, \cdot) is nondegenerate, and thus induces the desired isomorphism $\wedge^k(V^*) \xrightarrow{\cong} (\wedge^k(V))^*$. Doing this for each k shows that the exterior algebra

$$\wedge(V^*) = \bigoplus_{k=0}^{\dim V} \wedge^k(V^*) \cong \bigoplus_{k=0}^{\dim V} (\wedge^k(V))^* \cong \left(\bigoplus_{k=0}^{\dim V} \wedge^k(V) \right)^* = (\wedge(V))^*.$$

□

Notice that this argument—and especially (2.2)—really validates the idea that all alternating multilinear maps are, in some sense, determinants.

2.4.1 Differential Forms in Local Coordinates

[Definition 2.4.9](#) is not the easiest thing to compute with, so let's try to understand what differential forms look like in local coordinates. Suppose M is an n -manifold, $\omega \in \Omega^k(M)$ is a k -form, and $p \in M$. Let (U, ϕ) be a local coordinate chart in a neighborhood of p , and define the functions $x_i : \phi(U) \rightarrow \mathbb{R}$ by $x_i(\phi(a_1, \dots, a_n)) = a_i$. In other words, the function x_i reports the i th local coordinate of a point in this coordinate chart.

Now, each function x_i has a differential dx_i which (as in [Example 2.1.7](#)) is a 1-form on $\phi(U) \subset M$. In particular, at p the differential $(dx_i)_p : T_p M \rightarrow T_{x_i(p)} \mathbb{R} \cong \mathbb{R}$ is a linear functional on $T_p M$; that is, an element of the dual space $(T_p M)^*$.

If $\left\{ \frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_n} \right\}$ is the local coordinate basis for $T_p M$ associated to our chart, I claim that $\{(dx_1)_p, \dots, (dx_n)_p\}$ gives the corresponding dual basis for $(T_p M)^*$: by [Lemma 1.3.4](#),

$$(dx_i)_p \left(\frac{\partial}{\partial x_j} \right) = \frac{\partial}{\partial x_j}(x_i) = \delta_{ij}.$$

In what follows I'm going to drop the subscript p and just say $\{dx_1, \dots, dx_n\}$ is a basis for $(T_p M)^* = \wedge^1((T_p M)^*)$. More generally, as in [Proposition 2.4.5](#), $\{dx_{i_1} \wedge \cdots \wedge dx_{i_k} : 1 \leq i_1 < \cdots < i_k \leq n\}$ gives a basis for $\wedge^k((T_p M)^*)$, and hence at the point p our k -form ω can be written as

$$\omega_p = \sum_I a_I dx_I,$$

⁹At a more basic level, it's not hard to see that these vector spaces have the same dimension, and hence must be abstractly isomorphic, but it's helpful to make this isomorphism more concrete with the pairing.

where $I = (i_1, \dots, i_k)$ is any k -tuple of distinct, sorted indices and dx_I is shorthand for $dx_{i_1} \wedge \dots \wedge dx_{i_k}$.

Example 2.4.12. On \mathbb{R}^2 we have global coordinates, so every $\omega \in \Omega^1(\mathbb{R}^2)$ is of the form $\omega = f dx + g dy$ for smooth functions $f, g \in C^\infty(\mathbb{R}^2)$. So if $\omega = f dx + g dy$ and $\eta = h dx + m dy$ are two 1-forms, we will have

$$\omega \wedge \eta = (f dx + g dy) \wedge (h dx + m dy) = (fm - gh)dx \wedge dy.$$

In particular,

$$(\omega \wedge \eta)(U, V) = (\omega \wedge \eta) \left(u_1 \frac{\partial}{\partial x} + u_2 \frac{\partial}{\partial y}, v_1 \frac{\partial}{\partial x} + v_2 \frac{\partial}{\partial y} \right) = (fm - gh)(u_1 v_2 - u_2 v_1).$$

Example 2.4.13. On \mathbb{R}^3 , if $\omega, \eta \in \Omega^1(\mathbb{R}^3)$ with $\omega = f dx + g dy + h dz$ and $\eta = a dx + b dy + c dz$, then

$$\omega \wedge \eta = (gc - hb)dy \wedge dz + (ha - fc)dz \wedge dx + (fb - ga)dx \wedge dy.$$

Compare with the cross product

$$\omega^\sharp \times \eta^\sharp = \left(f \frac{\partial}{\partial x} + g \frac{\partial}{\partial y} + h \frac{\partial}{\partial z} \right) \times \left(a \frac{\partial}{\partial x} + b \frac{\partial}{\partial y} + c \frac{\partial}{\partial z} \right) = (gc - hb) \frac{\partial}{\partial x} + (ha - fc) \frac{\partial}{\partial y} + (fb - ga) \frac{\partial}{\partial z}.$$

Also, if $\mu = p dy \wedge dz + q dz \wedge dx + r dx \wedge dy$, then

$$\omega \wedge \mu = (f dx + g dy + h dz) \wedge (p dy \wedge dz + q dz \wedge dx + r dx \wedge dy) = (fp + gq + hr)dx \wedge dy \wedge dz.$$

Compare with the dot product

$$\omega^\sharp \cdot (\star \mu)^\sharp = \left(f \frac{\partial}{\partial x} + g \frac{\partial}{\partial y} + h \frac{\partial}{\partial z} \right) \times \left(p \frac{\partial}{\partial x} + q \frac{\partial}{\partial y} + r \frac{\partial}{\partial z} \right) = fp + gq + hr.$$

Locally, any manifold looks like \mathbb{R}^n , so these computations work just as well in local coordinates on 2- and 3-manifolds.

2.5 The Exterior Derivative

We've already seen some calculations of the exterior derivative in local coordinates in [Section 2.2](#), but to give a coordinate-free definition of the exterior derivative we need to talk about derivations.

Definition 2.5.1. If V is a vector space, an endomorphism $D : \wedge(V) \rightarrow \wedge(V)$ of the exterior algebra is

- a *derivation* if $D(u \wedge v) = (Du) \wedge v + u \wedge (Dv)$
- an *anti-derivation* if $D(u \wedge v) = (Du) \wedge v + (-1)^k u \wedge (Dv)$ for $u \in \wedge^k(V)$ (notice that this implies that $D(v_1 \wedge \dots \wedge v_m) = \sum_{i=1}^m (-1)^{i-1} v_1 \wedge \dots \wedge v_{i-1} \wedge (Dv_i) \wedge v_{i+1} \wedge \dots \wedge v_m$ for any $v_1, \dots, v_m \in V = \wedge^1(V)$)
- of *degree k* if $D : \wedge^i(V) \rightarrow \wedge^{i+k}(V)$ for all i .

(This definition extends to any (graded) algebra by substituting the product in the algebra for \wedge in the above.)

Example 2.5.2. We saw on page 8 that tangent vectors can be viewed as (ungraded) derivations on the algebra \mathcal{D}_p of functions which are smooth in a neighborhood of a point $p \in M$. More generally, a vector field $X \in \mathfrak{X}(M)$ is a derivation on the algebra $C^\infty(M)$ because it satisfies the Leibniz rule $X(fg) = (Xf)g + f(Xg)$.

There's no degree in either of these cases because \mathcal{D}_p and $C^\infty(M)$ are ungraded.

Definition 2.5.3. The *exterior derivative* $d : \Omega^*(M) \rightarrow \Omega^*(M)$ is the unique anti-derivation of degree +1 so that

- (i) For all $f \in C^\infty(M) = \Omega^0(M)$, df is the differential of f .
- (ii) $d \circ d = 0$.

Proposition 2.5.4. *There is such a creature and it is unique.*

The idea is that we kind of know what such an thing should look like in local coordinates around a point $p \in M$: any k -form $\omega \in \Omega^k(M)$ can be written in local coordinates as

$$\omega_p = \sum_I f_I dx_I.$$

So then define

$$(d\omega)_p := \sum_I (df_I) \wedge dx_I. \quad (2.3)$$

Example 2.5.5. Consider $\omega \in \Omega^1(\mathbb{R}^3)$ given by $\omega = z dy + xy dz$. Then

$$\begin{aligned} d\omega &= (dz) \wedge dy + (d(xy)) \wedge dz = dz \wedge dy + \left(\frac{\partial(xy)}{\partial x} dx + \frac{\partial(xy)}{\partial y} dy + \frac{\partial(xy)}{\partial z} dz \right) \wedge dz \\ &= dz \wedge dy + y dx \wedge dz + x dy \wedge dz = y dx \wedge dz + (x - 1)dy \wedge dz. \end{aligned}$$

I claim that (2.3) is the only way to get an anti-derivation of degree +1 that satisfies (i) and (ii) in local coordinates. Why? If d satisfies (i) and (ii), then $dx_i = d(x_i)$ is the result of applying d to the 0-form x_i (by (i)), so $d(dx_i) = (d \circ d)(x_i) = 0$ (by (ii)). But then the anti-derivation property implies that

$$d(dx_I) = d(dx_{i_1} \wedge \cdots \wedge dx_{i_k}) = \sum_{j=1}^k (-1)^{j-1} dx_{i_1} \wedge \cdots \wedge dx_{i_{j-1}} \wedge d(dx_{i_j}) \wedge dx_{i_{j+1}} \wedge \cdots \wedge dx_{i_k} = 0$$

since each $d(dx_{i_j}) = 0$. But then, again applying the anti-derivation condition together with linearity,

$$(d\omega)_p = d \left(\sum_I f_I dx_I \right) = \sum_I [(df_I) \wedge dx_I + f_I d(dx_I)] = \sum_I (df_I) \wedge dx_I.$$

Exercise 2.5.6. Go back to Section 2.2 and verify the various exterior derivative calculations we did there.

Proof of Proposition 2.5.4. Define d in local coordinates as in (2.3). Independence of the choice of coordinates will follow from uniqueness, so it remains to show:

- (a) This is an anti-derivation.

(b) It satisfies the cocycle condition (ii).

(c) Uniqueness.

(a) and (b) are both computations. Here's the computation for (b): If $\omega = \sum_I f_I dx_I$ in a neighborhood of $p \in M$, then

$$d\omega = \sum_I df_I \wedge dx_I = \sum_I \left(\sum_{i=1}^n \frac{\partial f}{\partial x_i} dx_i \right) \wedge dx_I$$

in that same neighborhood, so

$$d(d\omega) = \sum_I \left(\sum_{i=1}^n d \left(\frac{\partial f}{\partial x_i} \right) \wedge dx_i \right) \wedge dx_I = \sum_I \left(\sum_{i=1}^n \left(\sum_{j=1}^n \frac{\partial^2 f}{\partial x_j \partial x_i} dx_j \right) \wedge dx_i \right) \wedge dx_I = 0$$

since $\frac{\partial^2 f}{\partial x_j \partial x_i} = \frac{\partial^2 f}{\partial x_i \partial x_j}$ but $dx_j \wedge dx_i = -dx_i \wedge dx_j$.

Turning to (c), suppose there exists a degree +1 anti-derivation $D : \Omega^*(M) \rightarrow \Omega^*(M)$ satisfying (i) and (ii). In particular, $Df = df$ for all $f \in C^\infty(M) = \Omega^0(M)$. Then

$$D(dx_I) = D(dx_{i_1} \wedge \cdots \wedge dx_{i_k}) = \sum_{j=1}^k (-1)^{j-1} dx_{i_1} \wedge \cdots \wedge dx_{i_{j-1}} \wedge D(dx_{i_j}) \wedge dx_{i_{j+1}} \wedge \cdots \wedge dx_{i_k} = 0$$

since $D(dx_{i_j}) \stackrel{(i)}{=} D(Dx_{i_j}) \stackrel{(ii)}{=} 0$ by hypothesis.

Therefore, if $\omega = \sum_I f_I dx_I$ in a neighborhood of p , by linearity and the anti-derivation condition we have

$$D\omega = \sum_I (Df_I \wedge dx_I + f_I D(dx_I)) = \sum_I Df_I \wedge dx_I \stackrel{(i)}{=} \sum_I df_I \wedge dx_I = d\omega$$

in that neighborhood. Since this is true for all $p \in M$, it must be the case that $D\omega = d\omega$. \square

Notice that d being degree +1 and satisfying the cocycle condition (ii) says exactly that we have a diagram

$$\Omega^0(M) \xrightarrow{d} \Omega^1(M) \xrightarrow{d} \cdots \xrightarrow{d} \Omega^{n-1}(M) \xrightarrow{d} \Omega^n(M)$$

which is a cochain complex.

Definition 2.5.7. A differential form $\omega \in \Omega^k(M)$ is *closed* if $d\omega = 0$ and it is *exact* if $\omega = d\eta$ for some $\eta \in \Omega^{k-1}(M)$. Notice that, by the cocycle condition (ii), all exact forms are closed.

Exercise 2.5.8. Show that $\omega \in \Omega^1(\mathbb{R}^2 - \{0\})$ given by $\omega = \frac{x dy - y dx}{x^2 + y^2}$ is closed but not exact. (Using the correspondence between 1-forms and vector fields, this is equivalent to the fact that the vector field $\left(\frac{-y}{x^2 + y^2}, \frac{x}{x^2 + y^2} \right)$ is curl-free but not a gradient field.)

2.6 Pullbacks

Suppose we have manifolds M and N and a smooth map $f: M \rightarrow N$. Since we have a change of variables formula for integrals, and differential forms are supposed to be the objects that we can integrate on manifolds, we should expect that f gives a way of turning differential forms on one of M or N into forms on the other. If you think about how the change of variables formula works, you can figure out which direction this is supposed to go, but let's build up to this a bit more slowly.

First of all, recall that if $v \in T_p M$, then $df_p v \in T_{f(p)} N$; that is, $df_p: T_p M \rightarrow T_{f(p)} N$. More precisely, since $v = \alpha'(0)$ for some smooth curve α with $\alpha(0) = p$, we defined $df_p v = (f \circ \alpha)'(0)$. In words, f gives a way of pushing tangent vectors forward using df ,¹⁰ and we could have guessed this in advance because v is defined in terms of a map into M , which can then be composed with f to get a map into N . In categorical terms, tangent vectors and vector fields are transformed *covariantly*, meaning that tangent vectors or vector fields get sent in the same direction (from M to N) as the original map $f: M \rightarrow N$.

Now let's think about how we could move differential forms around using f . At a point, a k -form $\omega \in \Omega^k(M)$ corresponds to $\omega_p \in \wedge^k((T_p M)^*) \cong (\wedge^k(T_p M))^*$, or equivalently an alternating multilinear functional $(T_p M)^k \rightarrow \mathbb{R}$. In other words, unlike a tangent vector which is defined by a map *into* M , a differential form corresponds to a map *out of* $(T_p M)^k$. That means there's really no way to compose $\omega_p \in \wedge^k((T_p M)^*)$ with f to get something on N .

However, if we start instead with $\eta \in \Omega^k(N)$, then, as above, at a point $q \in N$, η_q corresponds to a map $(T_q N)^k \rightarrow \mathbb{R}$, so we *should* be able to precompose with a map $(T_p M)^k \rightarrow (T_q N)^k$ to get a k -form on M . More precisely:

Definition 2.6.1. If $f: M \rightarrow N$ is smooth and $\eta \in \Omega^k(N)$, then the *pullback* of η by f is the form $f^* \eta \in \Omega^k(M)$ defined by

$$(f^* \eta)_p(v_1, \dots, v_k) := \eta_{f(p)}(df_p v_1, \dots, df_p v_k)$$

for any $v_1, \dots, v_k \in T_p M$.

Notice, first of all, that this definition makes sense because $df_p: T_p M \rightarrow T_{f(p)} N$, so each $df_p v_i \in T_{f(p)} N$, and hence can be fed into $\eta_{f(p)}$.

In the paragraph before Definition 2.6.1 I said we could precompose η with a map $(T_p M)^k \rightarrow (T_q N)^k$, but then didn't write down such a map. What's going on here? The differential $df_p: T_p M \rightarrow T_{f(p)} N$ induces a map $(df_p)^k: (T_p M)^k \rightarrow (T_{f(p)} N)^k$ by just applying df_p separately in each factor. So then another way of stating Definition 2.6.1 is that

$$(f^* \eta)_p = \eta_{f(p)} \circ (df_p)^k.$$

Example 2.6.2. Suppose $g \in \Omega^0(N) = C^\infty(N)$. then

$$(f^* g)_p = g_{f(p)} = g(f(p)) = (g \circ f)(p),$$

recalling that, for a form ω , the notation ω_p really just means $\omega(p)$.

In other words, $f^* g = g \circ f$.

¹⁰In fact, many books and papers use the notation f_* rather than df since this is the standard notation for a pushforward operator.

Example 2.6.3. Let $f: S^1 \rightarrow \mathbb{R}^2$ be the inclusion of the unit circle into the plane; i.e., $f(\theta) = (\cos \theta, \sin \theta)$. Let $\omega = dx \in \Omega^1(\mathbb{R}^2)$, which at each point $p \in \mathbb{R}^2$ is computed by $dx \left(a \frac{\partial}{\partial x} + b \frac{\partial}{\partial y} \right) = a$; or, said another way, by taking the dot product of the input tangent vector with the first standard basis vector.

What is $f^*\omega = f^*dx$? At any point $\theta_0 \in S^1$, we have the local coordinate basis $\left\{ \frac{\partial}{\partial \theta} \right\}$ for $T_{\theta_0}S^1$, so we can completely specify f^*dx by determining the value of

$$(f^*dx)_{\theta_0} \left(a \frac{\partial}{\partial \theta} \right)$$

for any $a \in \mathbb{R}$. By Definition 2.6.1,

$$(f^*dx)_{\theta_0} \left(a \frac{\partial}{\partial \theta} \right) = (dx)_{f(\theta_0)} \left(df_{\theta_0} \left(a \frac{\partial}{\partial \theta} \right) \right).$$

So first we'll need to compute $df_{\theta_0} \left(a \frac{\partial}{\partial \theta} \right)$. We can write $a \frac{\partial}{\partial \theta} = \alpha'(0)$, where $\alpha(t) = \theta_0 + at$, so Definition 1.3.1 tells us that

$$\begin{aligned} df_{\theta_0} \left(a \frac{\partial}{\partial \theta} \right) &= (f \circ \alpha)'(0) = \frac{d}{dt} \Big|_{t=0} f(\alpha(t)) = \frac{d}{dt} \Big|_{t=0} f(\theta_0 + at) = \frac{d}{dt} \Big|_{t=0} (\cos(\theta_0 + at), \sin(\theta_0 + at)) \\ &= (-a \sin(\theta_0 + at), a \cos(\theta_0 + at))|_{t=0} = (-a \sin(\theta_0), a \cos(\theta_0)) = -a \sin(\theta_0) \frac{\partial}{\partial x} + a \cos(\theta_0) \frac{\partial}{\partial y} \end{aligned}$$

after switching between tuple notation and local coordinate basis notation.

Therefore,

$$(f^*dx)_{\theta_0} \left(a \frac{\partial}{\partial \theta} \right) = (dx)_{f(\theta_0)} \left(df_{\theta_0} \left(a \frac{\partial}{\partial \theta} \right) \right) = (dx)_{f(\theta_0)} \left(-a \sin(\theta_0) \frac{\partial}{\partial x} + a \cos(\theta_0) \frac{\partial}{\partial y} \right) = -a \sin(\theta_0).$$

In other words,

$$(f^*dx)_{\theta_0} = -\sin \theta_0 (d\theta)_{\theta_0}$$

or, by slight abuse of notation,

$$f^*dx = -\sin \theta d\theta.$$

Recalling that $\theta \in S^1$ maps to the point $(\cos \theta, \sin \theta)$, we can see that $\sin \theta$ is the y -coordinate, so this could also be written as

$$f^*dx = -y d\theta.$$

Geometrically, this makes sense: at the point $(x, y) = (\cos \theta, \sin \theta)$ this form is dual to the tangent vector $-y \frac{\partial}{\partial \theta}$, which at each point is just the orthogonal projection of the first standard basis vector $\frac{\partial}{\partial x}$ onto the tangent space to the circle, and of course $\frac{\partial}{\partial x}$ is the tangent vector to $(x, y) \in \mathbb{R}^2$ dual to dx . See Figure 2.2.

One nice thing about pullbacks is that they commute with the exterior derivative:

Proposition 2.6.4. For $f: M \rightarrow N$ smooth and $\omega \in \Omega^k(N)$,

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

Before proving this, let's see it in an example:

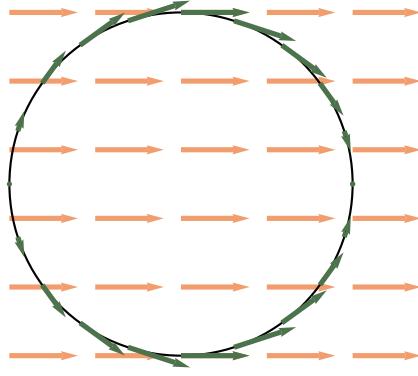


Figure 2.2: The vector field $\frac{\partial}{\partial x}$ (dual to dx) on \mathbb{R}^2 is shown in orange, and the vector field $-y \frac{\partial}{\partial \theta}$ (dual to $-y d\theta = f^* dx$) on the circle is shown in green.

Example 2.6.5. Continuing Example 2.5.5, recall that $\omega = z dy + xy dz$ and we computed

$$d\omega = y dx \wedge dz + (x - 1)dy \wedge dz.$$

Now, consider the map $f: \mathbb{R}^2 \rightarrow \mathbb{R}^3$ given by $f(u, v) = (u^2, u, uv)$. For $p = (u_0, v_0) \in \mathbb{R}^2$ and $w = a \frac{\partial}{\partial u} + b \frac{\partial}{\partial v} \in T_p \mathbb{R}^2$ for $i = 1, 2$, we can write $w = \alpha'_i(0)$ where $\alpha_i(t) = p + tw = (u_0, v_0) + t(a, b)$, so

$$\begin{aligned} df_p(w) &= (f \circ \alpha_i)'(0) = \left. \frac{d}{dt} \right|_{t=0} ((u_0 + ta)^2, u_0 + ta, (u_0 + ta)(v_0 + tb)) \\ &= (2au_0, a, av_0 + bu_0) = 2au_0 \frac{\partial}{\partial x} + a \frac{\partial}{\partial y} + (av_0 + bu_0) \frac{\partial}{\partial z}, \end{aligned}$$

where I'm going back and forth between tuple notation (in \mathbb{R}^3 with the standard basis) and local coordinate basis notation (on $T_{f(p)} \mathbb{R}^3$ with the local coordinate basis [which of course is the same as the standard basis]).

Therefore,

$$\begin{aligned} (f^* \omega)_p(w) &= \omega_{f(p)}(df_p(w)) = \left(u_0 v_0 dy + u_0^3 dz \right) \left(2au_0 \frac{\partial}{\partial x} + a \frac{\partial}{\partial y} + (av_0 + bu_0) \frac{\partial}{\partial z} \right) \\ &= au_0 v_0 (1 + u_0^2) + bu_0^4. \end{aligned}$$

Since this is what the linear functional $(f^* \omega)_{(u_0, v_0)}$ does to the tangent vector $w = a \frac{\partial}{\partial u} + b \frac{\partial}{\partial v}$, we see that $(f^* \omega)_{(u_0, v_0)} = u_0 v_0 (1 + u_0^2) du + u_0^4 dv$ or, now thinking globally,

$$f^* \omega = uv(1 + u^2)du + u^4dv.$$

Hence,

$$\begin{aligned}
d(f^*\omega) &= \left(d(uv(1+u^2)) \right) \wedge du + \left(d(u^4) \right) \wedge dv \\
&= \left(\frac{\partial}{\partial u} (uv(1+u^2)) du + \frac{\partial}{\partial v} (uv(1+u^2)) dv \right) \wedge du + \left(\frac{\partial}{\partial u} (u^4) du + \frac{\partial}{\partial v} (u^4) dv \right) \wedge dv \\
&= u(1+u^2)dv \wedge du + 4u^3du \wedge dv \\
&= (3u^3 - u)du \wedge dv.
\end{aligned}$$

On the other hand,

$$\begin{aligned}
(f^*d\omega)_p \left(\frac{\partial}{\partial u}, \frac{\partial}{\partial v} \right) &= (d\omega)_{f(p)} \left(df_p \left(\frac{\partial}{\partial u} \right), df_p \left(\frac{\partial}{\partial v} \right) \right) \\
&= \left(u_0 dx \wedge dz + (u_0^2 - 1)dy \wedge dz \right) \left(2u_0 \frac{\partial}{\partial x} + \frac{\partial}{\partial y} + v_0 \frac{\partial}{\partial z}, u_0 \frac{\partial}{\partial z} \right) \\
&= 3u_0^3 - u_0.
\end{aligned}$$

Notice that $(3u_0^3 - u_0)du \wedge dv$ would produce the same output for this (or any other) input, so $(f^*d\omega)_{(u_0, v_0)} = (3u_0^3 - u_0)du \wedge dv$. In other words (now thinking globally)

$$f^*d\omega = (3u^3 - u)du \wedge dv = d(f^*\omega),$$

as predicted by [Proposition 2.6.4](#).

Proof of Proposition 2.6.4. Let $p \in M$. If $\omega = g \in \Omega^0(N) = C^\infty(N)$, then $f^*g = g \circ f$, so $d(f^*g) = d(g \circ f)$. On the other hand, for any $v \in T_p M$,

$$(f^*dg)_p v = dg_{f(p)}(df_p v) = d(g \circ f)_p(v) = d(f^*g)_p(v),$$

where we used [Definition 2.6.1](#) for the first equality, the Chain Rule for the second, and [Example 2.6.2](#) for the third. This completes the proof in the $k = 0$ case.

Now, suppose $k > 0$ and that x_1, \dots, x_n are local coordinates in a neighborhood of $f(p) \in N$, meaning that $\omega \in \Omega^k(N)$ has the form $\omega_{f(p)} = \sum_I g_I dx_I$ at $f(p)$, so that

$$(d\omega)_{f(p)} = \sum_I dg_I \wedge dx_I.$$

But then

$$(f^*d\omega)_p = \sum_I (f^*dg_I) \wedge (f^*dx_I) = \sum_I d(g_I \circ f) \wedge f^*dx_{i_1} \wedge \cdots \wedge f^*dx_{i_k} = \sum_I d(g_I \circ f) \wedge d(x_{i_1} \circ f) \wedge \cdots \wedge d(x_{i_k} \circ f),$$

where we've repeatedly used both the $k = 0$ case and the fact that $f^*(\alpha \wedge \beta) = (f^*\alpha) \wedge (f^*\beta)$ (which follows from HW #2, Problem 1).

On the other hand,

$$(f^*\omega)_p = \sum_I (g_I \circ f) f^*dx_I = \sum_I (g_I \circ f) f^*dx_{i_1} \wedge \cdots \wedge f^*dx_{i_k} = \sum_I (g_I \circ f) d(x_{i_1} \circ f) \wedge \cdots \wedge d(x_{i_k} \circ f),$$

so

$$(df^*\omega)_p = \sum_I d(g_I \circ f) \wedge d(x_{i_1} \circ f) \wedge \cdots \wedge d(x_{i_k} \circ f) = (f^*d\omega)_p,$$

and we conclude that the proposition is also true for $k > 0$. \square

2.7 Lie Derivatives and Cartan's Magic Formula

Recall [Definition 1.7.4](#), the definition of the Lie derivative for vector fields:

$$\mathcal{L}_X Y := \lim_{t \rightarrow 0} \frac{(d\Phi_{-t})_{\Phi_t(p)} Y(\Phi_t(p)) - Y(p)}{t},$$

where Φ_t is the local flow for $X \in \mathfrak{X}(M)$. We saw in [Proposition 1.7.5](#) that the Lie derivative is equal to the Lie bracket $[X, Y]$.

The key idea in this derivative was to compare Y at $\Phi_t(p)$ to Y at p , and to do so we needed to move $Y(\Phi_t(p))$ from $T_{\Phi_t(p)} M$ to $T_p M$, which we did with the differential of the *negative* flow $d\Phi_{-t}$.

We can use the same idea to get a Lie derivative for differential forms, though we have to be careful about how we move a form $\omega \in \Omega^k(M)$ at a point $\Phi_t(p)$ to p : whereas vector fields push forward, differential forms pull back (as we've just seen in [Section 2.6](#)). So rather than pushing forward with $d\Phi_{-t}$, we will pull back by Φ_t^* :

Definition 2.7.1. Let $X \in \mathfrak{X}(M)$ and $\omega \in \Omega^k(M)$. Then the *Lie derivative* of ω with respect to X is defined (at a point $p \in M$) by

$$(\mathcal{L}_X \omega)_p := \lim_{t \rightarrow 0} \frac{\Phi_t^*(\omega_{\Phi_t(p)}) - \omega_p}{t} = \frac{d}{dt} \Big|_{t=0} (\Phi_t^*)_p (\omega_{\Phi_t(p)}).$$

Even more generally, we can combine pushforwards and pullbacks to get Lie derivatives of arbitrary tensor fields:

Definition 2.7.2. Let $X \in \mathfrak{X}(M)$, $a \in \mathcal{T}_r^0(M)$, and $\beta \in \mathcal{T}_0^s(M)$. Then the *Lie derivative* of the (r, s) -tensor field $a \otimes \beta$ is defined (at a point $p \in M$) by

$$(\mathcal{L}_X a \otimes \beta)_p := \frac{d}{dt} \Big|_{t=0} \left[(d\Phi_{-t})_{\Phi_t(p)} (a_{\Phi_t(p)}) \otimes (\Phi_t^*)_p (\beta_{\Phi_t(p)}) \right].$$

Extending linearly gives the Lie derivative with respect to X for arbitrary elements of $\mathcal{T}_r^s(M)$.

Lemma 2.7.3. For $f \in C^\infty(M) = \Omega^0(M)$ and $X \in \mathfrak{X}(M)$,

$$\mathcal{L}_X f = X(f).$$

Proof. By [Definition 2.7.1](#) (first equality), [Example 2.6.2](#) (second), and [Definition 1.2.1](#) (third),

$$\mathcal{L}_X f = \frac{d}{dt} \Big|_{t=0} \Phi_t^*(f) = \frac{d}{dt} \Big|_{t=0} f \circ \Phi_t = X(f).$$

□

Example 2.7.4. If M is a manifold and g is a Riemannian metric on M , a vector field $X \in \mathfrak{X}(M)$ is called a *Killing field*¹¹ if $\mathcal{L}_X g = 0$. Intuitively, this means that the flow generated by X is an isometry: it preserves distances. If you read about general relativity, you will often see references to Killing fields, which encode the symmetries of spacetime.

¹¹Named after Wilhelm Killing.

Let's compute $\mathcal{L}_X g$ for some vector fields X on S^2 , where g is the standard Riemannian metric induced by the inner product on \mathbb{R}^3 . If we work in cylindrical coordinates, we have the local coordinate chart $\phi: (0, 2\pi) \rightarrow S^2$ given by

$$\phi(\theta, z) = \left(\sqrt{1 - z^2} \cos \theta, \sqrt{1 - z^2} \sin \theta, z \right).$$

So then our local coordinate basis $\{\frac{\partial}{\partial \theta}, \frac{\partial}{\partial z}\}$ at a point $p = \phi(\theta_0, z_0)$ looks like

$$\frac{\partial}{\partial \theta} \Big|_p = \frac{d}{dt} \Big|_{t=0} \phi(\theta_0 + t, z_0) = \left(-\sqrt{1 - z_0^2} \sin \theta_0, \sqrt{1 - z_0^2} \cos \theta_0, z_0 \right)$$

and

$$\frac{\partial}{\partial z} \Big|_p = \frac{d}{dt} \Big|_{t=0} \phi(\theta_0, z_0 + t) = \left(-\frac{z_0 \cos \theta_0}{\sqrt{1 - z_0^2}}, -\frac{z_0 \sin \theta_0}{\sqrt{1 - z_0^2}}, 1 \right).$$

See [Figure 2.3](#) (which is the same as [Figure 1.11](#)).

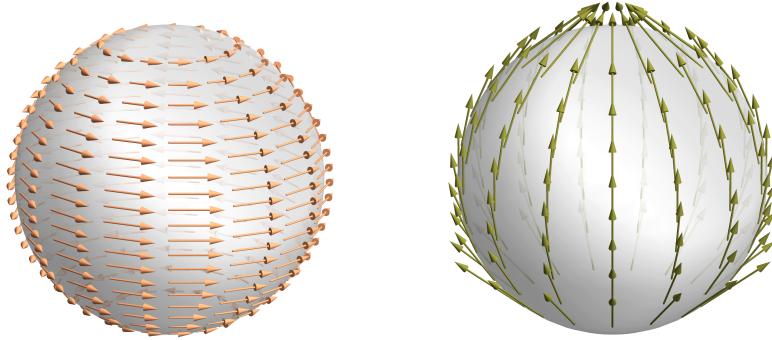


Figure 2.3: The vector fields $X = \frac{\partial}{\partial \theta}$ and $Y = \frac{\partial}{\partial z}$ on S^2 .

In turn, this implies that the Riemannian metric g is given by

$$\begin{aligned} g_p \left(a \frac{\partial}{\partial \theta} + b \frac{\partial}{\partial z}, c \frac{\partial}{\partial \theta} + d \frac{\partial}{\partial z} \right) &= \left[a \left(-\sqrt{1 - z_0^2} \sin \theta_0, \sqrt{1 - z_0^2} \cos \theta_0, z_0 \right) + b \left(-\frac{z_0 \cos \theta_0}{\sqrt{1 - z_0^2}}, -\frac{z_0 \sin \theta_0}{\sqrt{1 - z_0^2}}, 1 \right) \right. \\ &\quad \cdot \left. \left[c \left(-\sqrt{1 - z_0^2} \sin \theta_0, \sqrt{1 - z_0^2} \cos \theta_0, z_0 \right) + d \left(-\frac{z_0 \cos \theta_0}{\sqrt{1 - z_0^2}}, -\frac{z_0 \sin \theta_0}{\sqrt{1 - z_0^2}}, 1 \right) \right] \right] \\ &= (1 - z_0^2) ac + \frac{1}{1 - z_0^2} bd \end{aligned}$$

or, more symbolically,

$$g_{(\theta,z)} = (1 - z^2) d\theta \otimes d\theta + \frac{1}{1 - z^2} dz \otimes dz.$$

Notice that this only depends on z and not on θ , so we would expect that flowing by $\frac{\partial}{\partial \theta}$ should preserve this inner product; that is, $\frac{\partial}{\partial \theta}$ should be a Killing field.

Indeed, the local flow of $\frac{\partial}{\partial \theta}$ is given at a point $p = \phi(\theta_0, z_0)$ by

$$\Phi_t(p) = \phi(\theta_0 + t, z) = \left(\sqrt{1 - z_0^2} \cos(\theta_0 + t), \sqrt{1 - z_0^2} \sin(\theta_0 + t), z_0 \right),$$

Now, writing $\frac{\partial}{\partial \theta} = \alpha'(0)$ where $\alpha(s) = \phi(\theta_0 + s, z_0)$, we see that

$$\begin{aligned} (d\Phi_t)_p \left(\frac{\partial}{\partial \theta} \Big|_p \right) &= (d\Phi_t)_p \alpha'(0) = \frac{d}{ds} \Big|_{s=0} (\Phi_t \circ \alpha)(s) \\ &= \frac{d}{ds} \Big|_{s=0} (\Phi_t(\phi(\theta_0 + s, z_0))) \\ &= \frac{d}{ds} \Big|_{s=0} \left(\sqrt{1 - z_0^2} \cos(\theta_0 + s + t), \sqrt{1 - z_0^2} \sin(\theta_0 + s + t), z_0 \right) \\ &= \left(-\sqrt{1 - z_0^2} \sin(\theta_0 + t), \sqrt{1 - z_0^2} \cos(\theta_0 + t), 0 \right) \\ &= \frac{\partial}{\partial \theta} \Big|_{\Phi_t(p)}; \end{aligned}$$

i.e., the local flow of $\frac{\partial}{\partial \theta}$ preserves $\frac{\partial}{\partial \theta}$, as you might expect.

On the other hand, $\frac{\partial}{\partial z} = \beta'(0)$ where $\beta(s) = \phi(\theta_0, z_0 + s)$, so

$$\begin{aligned} (d\Phi_t)_p \left(\frac{\partial}{\partial z} \Big|_p \right) &= (d\Phi_t)_p \beta'(0) = \frac{d}{ds} \Big|_{s=0} (\Phi_t \circ \beta)(s) \\ &= \frac{d}{ds} \Big|_{s=0} (\Phi_t(\phi(\theta_0, z_0 + s))) \\ &= \frac{d}{ds} \Big|_{s=0} \left(\sqrt{1 - (z_0 + s)^2} \cos(\theta_0 + t), \sqrt{1 - (z_0 + s)^2} \sin(\theta_0 + t), z_0 + s \right) \\ &= \left(-\frac{z_0 \cos(\theta_0 + t)}{\sqrt{1 - z_0^2}}, -\frac{z_0 \sin(\theta_0 + t)}{\sqrt{1 - z_0^2}}, 1 \right) \\ &= \frac{\partial}{\partial z} \Big|_{\Phi_t(p)}, \end{aligned}$$

so flowing by $\frac{\partial}{\partial \theta}$ also preserves $\frac{\partial}{\partial z}$.

Therefore, for any $a \frac{\partial}{\partial \theta} \Big|_p + b \frac{\partial}{\partial z} \Big|_p \in T_p S^2$, we have

$$(d\Phi_t)_p \left(a \frac{\partial}{\partial \theta} \Big|_p + b \frac{\partial}{\partial z} \Big|_p \right) = a \frac{\partial}{\partial \theta} \Big|_{\Phi_t(p)} + b \frac{\partial}{\partial z} \Big|_{\Phi_t(p)}.$$

Since p and $\Phi_t(p)$ have the same z -coordinate,

$$\begin{aligned} g_{\Phi_t(P)} \left(a \frac{\partial}{\partial \theta} \Big|_{\Phi_t(P)} + b \frac{\partial}{\partial z} \Big|_{\Phi_t(P)}, c \frac{\partial}{\partial \theta} \Big|_{\Phi_t(P)} + d \frac{\partial}{\partial z} \Big|_{\Phi_t(P)} \right) &= (1 - z_0^2)ac + \frac{1}{1 - z_0^2}bd \\ &= g_p \left(a \frac{\partial}{\partial \theta} \Big|_p + b \frac{\partial}{\partial z} \Big|_p, c \frac{\partial}{\partial \theta} \Big|_p + d \frac{\partial}{\partial z} \Big|_p \right). \end{aligned}$$

Hence, for any $u, v \in T_p S^2$

$$(\Phi_t^* g_{\Phi_t(P)})(u, v) = g_{\Phi_t(P)}((d\Phi_t)_p u, (d\Phi_t)_p v) = g_p(u, v)$$

is independent of t , and we see that

$$(\mathcal{L}_{\frac{\partial}{\partial \theta}} g)_p(u, v) = \frac{d}{dt} \Big|_{t=0} (\Phi_t^* g_{\Phi_t(P)})(u, v) = 0.$$

Therefore, $\frac{\partial}{\partial \theta}$ is a Killing field on the sphere, which makes sense: the local flow is just rotation around the z -axis, which certainly preserves distances.

On the other hand,

$$\Psi_t(p) = \phi(\theta_0, z_0 + t) = \left(\sqrt{1 - (z_0 + t)^2} \cos \theta_0, \sqrt{1 - (z_0 + t)^2} \sin \theta_0, z_0 + t \right)$$

is the local flow of $\frac{\partial}{\partial z}$. So then

$$\begin{aligned} (d\Psi_t)_p \left(\frac{\partial}{\partial \theta} \Big|_p \right) &= (d\Psi_t)_p \alpha'(0) = \frac{d}{ds} \Big|_{s=0} (\Psi_t \circ \alpha)(s) \\ &= \frac{d}{ds} \Big|_{s=0} (\Psi_t(\phi(\theta_0 + s, z_0))) \\ &= \frac{d}{ds} \Big|_{s=0} \left(\sqrt{1 - (z_0 + t)^2} \cos(\theta_0 + s), \sqrt{1 - (z_0 + t)^2} \sin(\theta_0 + s), z_0 + t \right) \\ &= \left(-\sqrt{1 - (z_0 + t)^2} \sin \theta_0, \sqrt{1 - (z_0 + t)^2} \cos \theta_0, 0 \right) \\ &= \frac{\partial}{\partial \theta} \Big|_{\Psi_t(p)} \end{aligned}$$

and

$$\begin{aligned}
(d\Psi_t)_P \left(\frac{\partial}{\partial z} \Big|_p \right) &= (d\Psi_t)_P \beta'(0) = \frac{d}{ds} \Big|_{s=0} (\Psi_t \circ \beta)(s) \\
&= \frac{d}{ds} \Big|_{s=0} (\Psi_t(\phi(\theta_0, z_0 + s))) \\
&= \frac{d}{ds} \Big|_{s=0} \left(\sqrt{1 - (z_0 + s + t)^2} \cos \theta_0, \sqrt{1 - (z_0 + s + t)^2} \sin \theta_0, z_0 + s + t \right) \\
&= \left(-\frac{(z_0 + t) \cos \theta_0}{\sqrt{1 - (z_0 + t)^2}}, -\frac{(z_0 + t) \sin \theta_0}{\sqrt{1 - (z_0 + t)^2}}, 1 \right) \\
&= \frac{\partial}{\partial z} \Big|_{\Psi_t(p)},
\end{aligned}$$

so flowing by $\frac{\partial}{\partial z}$ also preserves the coordinate vector fields, but the coordinate vector fields do not have constant length on S^2 .

We see that

$$\begin{aligned}
(\Psi_t^* g_{\Psi_t(p)}) \left(a \frac{\partial}{\partial \theta} \Big|_p + b \frac{\partial}{\partial z} \Big|_p, c \frac{\partial}{\partial \theta} \Big|_p + d \frac{\partial}{\partial z} \Big|_p \right) &= g_{\Psi_t(p)} \left(a \frac{\partial}{\partial \theta} \Big|_{\Psi_t(p)} + b \frac{\partial}{\partial z} \Big|_{\Psi_t(p)}, c \frac{\partial}{\partial \theta} \Big|_{\Psi_t(p)} + d \frac{\partial}{\partial z} \Big|_{\Psi_t(p)} \right) \\
&= (1 - (z_0 + t)^2)ac + \frac{1}{1 - (z_0 + t)^2}bd,
\end{aligned}$$

and hence

$$\begin{aligned}
(\mathcal{L}_{\frac{\partial}{\partial z}} g)_p \left(\frac{\partial}{\partial \theta} \Big|_p, \frac{\partial}{\partial \theta} \Big|_p \right) &= \frac{d}{dt} \Big|_{t=0} (\Psi_t^* g_{\Psi_t(p)}) \left(\frac{\partial}{\partial \theta} \Big|_p, \frac{\partial}{\partial \theta} \Big|_p \right) = \frac{d}{dt} \Big|_{t=0} \left[(1 - (z_0 + t)^2)ac + \frac{1}{1 - (z_0 + t)^2}bd \right] \\
&= -2z_0 ac + \frac{2z_0}{(1 - z_0^2)^2} bd;
\end{aligned}$$

that is,

$$\mathcal{L}_{\frac{\partial}{\partial z}} g = -2z d\theta \otimes d\theta + \frac{2z}{(1 - z^2)^2} dz \otimes dz,$$

which is no longer positive-definite (and hence not a Riemannian metric), but is still a symmetric $(0, 2)$ -tensor field on S^2 which captures how g changes as we flow in the $\frac{\partial}{\partial z}$ direction.

One of the most important formulas in differential geometry is:

Theorem 2.7.5 (Cartan's Magic Formula). *If $\omega \in \Omega^k(M)$ and $X \in \mathfrak{X}(M)$, then*

$$\mathcal{L}_X \omega = \iota_X d\omega + d\iota_X \omega.$$

Here ι_X is the *contraction operator* (or *interior product*) defined as follows: for $\omega \in \Omega^k(M)$ and $X \in \mathfrak{X}(M)$, $\iota_X \omega$ is the $(k - 1)$ -form defined by

$$(\iota_X \omega)_p(v_1, \dots, v_{k-1}) = \omega_p(X(p), v_1, \dots, v_{k-1})$$

for any $v_1, \dots, v_{k-1} \in T_p M$. This is sometimes also denoted by $X \lrcorner \omega$. This is an antiderivation of degree -1 .

The reason Cartan's magic formula is so important is that it is one of the most useful tools out there for computing exterior derivatives.

Example 2.7.6. Remember [Section 1.8](#), where we computed Lie brackets on S^3 . As a reminder, we had mutually perpendicular vector fields $X, Y, Z \in \mathfrak{X}(S^3)$ defined by

$$X(p) = pi, \quad Y(p) = pj, \quad Z(p) = pk$$

for all $p \in S^3$ thought of as unit quaternions, and showed that

$$[X, Y] = 2Z, \quad [Y, Z] = 2X, \quad [Z, X] = 2Y.$$

Let α, β, γ be the dual 1-forms; that is, $\alpha(X) = 1, \alpha(Y) = 0 = \alpha(Z)$ and similarly for β and γ . Then, using Cartan's magic formula,

$$d\alpha(Y, Z) = (\iota_Y d\alpha)(Z) = (\mathcal{L}_Y \alpha)(Z) - d(\iota_Y \alpha)(Z) = (\mathcal{L}_Y \alpha)(Z) - d(\alpha(Y))(Z) = (\mathcal{L}_Y \alpha)(Z)$$

since $\alpha(Y) = 0$.

The Lie derivative turns out to be a derivation, so it obeys a Leibniz rule:

$$\mathcal{L}_U(\omega(Y_1, \dots, Y_k)) = (\mathcal{L}_U \omega)(Y_1, \dots, Y_k) + \sum_{j=1}^k \omega(Y_1, \dots, \mathcal{L}_U Y_j, \dots, Y_k).$$

In particular,

$$0 = \mathcal{L}_Y(0) = \mathcal{L}_Y(\alpha(Z)) = (\mathcal{L}_Y \alpha)(Z) + \alpha(\mathcal{L}_Y Z) = (\mathcal{L}_Y \alpha)(Z) + \alpha([Y, Z]) = (\mathcal{L}_Y \alpha)(Z) + \alpha(2X) = (\mathcal{L}_Y \alpha)(Z) + 2,$$

so we conclude that

$$d\alpha(Y, Z) = (\mathcal{L}_Y \alpha)(Z) = -2.$$

By analogous calculations, we can show that $d\alpha(X, Y) = 0 = d\alpha(Z, X)$, so we conclude that

$$d\alpha = -2\beta \wedge \gamma,$$

and similarly one can show that $d\beta = 2\gamma \wedge \alpha$ and $d\gamma = -2\alpha \wedge \beta$.

This implies, in particular, that $\alpha \wedge d\alpha = -2\alpha \wedge \beta \wedge \gamma$ is never zero, so α is an example of what is called a *contact form* on S^3 .

Notice, in particular, that we never had to work in local coordinates in [Example 2.7.6](#), which is one of the great virtues of Cartan's magic formula: it is one of the few tools that allows one to compute without using coordinates.

Hopefully this gives you some reason to believe that Cartan's magic formula is useful, so let's try to prove it:

Proof of Theorem 2.7.5. If $u \in \Omega^0(M) = C^\infty(M)$, then $\mathcal{L}_X u = X(u)$ by [Lemma 2.7.3](#) and

$$\iota_X du + d\iota_X u = du(X) + 0 = du(X) = X(u) = \mathcal{L}_X u,$$

so the formula holds for 0-forms.

Now consider a 1-form ω which is the differential of a 0-form: $\omega = du$. If $Y \in \mathfrak{X}(M)$, then, by Definition 2.7.1,

$$(\mathcal{L}_X\omega)_p(Y) = (\mathcal{L}_x du)_p(Y) = \frac{d}{dt} \Big|_{t=0} \Phi_t^*(du_{\Phi_t(p)})(Y) = (\mathcal{L}_x du)_p(Y) = \frac{d}{dt} \Big|_{t=0} du_{\Phi_t(p)}((d\Phi_t)_p(Y))$$

using the definition of pullback (Definition 2.6.1). In turn, the chain rule tells us that the right hand side is equal to

$$\frac{d}{dt} \Big|_{t=0} d(u \circ \Phi_t)_p(Y) = \frac{d}{dt} \Big|_{t=0} Y(u \circ \Phi_t)(p)$$

by Lemma 1.3.4. Since Y is independent of t , the (1-dimensional) chain rule tell us that this is equal to

$$Y \left(\frac{d(u \circ \Phi_t)}{dt} \Big|_{t=0} \right) (p) = Y(X(u))(p).$$

So we have shown that $(\mathcal{L}_X\omega)_p(Y) = Y(X(u))(p)$.

On the other hand, $\iota_X d\omega = \iota_X d(du) = 0$ and

$$d(\iota_X\omega)(Y) = d(\iota_X du)(Y) = d(du(X))(Y) = d(X(U))(Y) = Y(X(U)) = (\mathcal{L}_X\omega)_p(Y)$$

by repeatedly using Lemma 1.3.4. Hence, the formula holds for differentials.

Finally, then, if $\omega \in \Omega^k(M)$ and we write $\omega = \sum_I a_I dx_I$ in local coordinates, then

$$\mathcal{L}_X\omega = \sum_I [(\mathcal{L}_X a_I) dx_I + a_I \mathcal{L}_X(dx_I)]. \quad (2.4)$$

The $k = 0$ argument above implies that

$$\mathcal{L}_X a_I = (\iota_X d + d\iota_X) a_I. \quad (2.5)$$

On the other hand, again using the fact that the Lie derivative is a derivation, we have

$$\begin{aligned} \mathcal{L}_X(dx_{i_1} \wedge \cdots \wedge dx_{i_k}) &= \sum_{j=1}^k dx_{i_1} \wedge \cdots \wedge \mathcal{L}_X dx_{i_j} \wedge \cdots \wedge dx_{i_k} = \sum_{j=1}^k dx_{i_1} \wedge \cdots \wedge (\iota_X d + d\iota_X) dx_{i_j} \wedge \cdots \wedge dx_{i_k} \\ &= (\iota_X d + d\iota_X)(dx_{i_1} \wedge \cdots \wedge dx_{i_k}) \end{aligned}$$

since each dx_{i_j} is a differential.

Combining this with (2.4) and (2.5) implies that

$$\mathcal{L}_X\omega = \sum_I [((\iota_X d + d\iota_X)a_I) dx_I + a_I(\iota_X d + d\iota_X)(dx_I)] = (\iota_X d + d\iota_X) \sum_I a_I dx_I = (\iota_X d + d\iota_X)\omega,$$

as desired. \square

2.8 De Rham Cohomology

Recall that we have the cochain complex

$$\Omega^0(M) \xrightarrow{d_0} \Omega^1(M) \xrightarrow{d_1} \dots \xrightarrow{d_{n-2}} \Omega^{n-1}(M) \xrightarrow{d_{n-1}} \Omega^n(M)$$

where the subscript k indicates that d_k is the restriction of d to $\Omega^k(M)$.

Definition 2.8.1. The k th de Rham cohomology group of M is

$$H_{\text{dR}}^k(M) := \frac{\ker d_k}{\text{im } d_{k-1}}.$$

Example 2.8.2. We saw in Exercise 2.5.8 that the 1-form $\omega = \frac{x \, dy - y \, dx}{x^2 + y^2} \in \Omega^1(\mathbb{R}^2 - \{0\})$ is *closed* (in the kernel of d), but not *exact* (in the image of d), so it represents a non-trivial element of $H_{\text{dR}}^1(\mathbb{R}^2 - \{0\})$.

De Rham cohomology is a homotopy invariant of manifolds, meaning that homotopy equivalent manifolds have identical de Rham cohomology groups.

If $\omega \in \Omega^k(M)$ and $\eta \in \Omega^\ell(M)$ are closed (i.e., $d\omega = 0$ and $d\eta = 0$), meaning that they represent de Rham cohomology classes in $H_{\text{dR}}^k(M)$ and $H_{\text{dR}}^\ell(M)$, respectively, then

$$d(\omega \wedge \eta) = d\omega \wedge \eta + (-1)^k \omega \wedge d\eta = 0 \wedge \eta + (-1)^k \omega \wedge 0 = 0,$$

so $d(\omega \wedge \eta)$ is also closed, and hence represents a class in $H_{\text{dR}}^{k+\ell}(M)$. This means that

$$H_{\text{dR}}^*(M) := \bigoplus_k H_{\text{dR}}^k(M)$$

forms a (graded) ring, the *de Rham cohomology ring*, with the operation induced by \wedge .

Theorem 2.8.3 (de Rham theorem). *There exists a vector space isomorphism*

$$H_{\text{dR}}^k(M) \rightarrow H_k(M; \mathbb{R})^* \cong H^k(M; \mathbb{R}),$$

where $H_k(M; \mathbb{R})$ and $H^k(M; \mathbb{R})$ are the singular (or simplicial) homology and cohomology groups.

In fact, even more is true: the de Rham cohomology ring is isomorphic to the singular cohomology ring with real coefficients.

In Example 2.8.2, $H^1(\mathbb{R}^2 - \{0\}; \mathbb{R}) \cong H^1(S^1; \mathbb{R}) \cong \mathbb{R}$, so the 1-form $\omega = \frac{x \, dy - y \, dx}{x^2 + y^2} \in \Omega^1(\mathbb{R}^2 - \{0\})$ is a generator of $H_{\text{dR}}^1(\mathbb{R}^2 - \{0\})$.

We won't give all the details of the proof of de Rham's theorem, but the basic idea is that a closed form defines a linear functional on homology classes by integration. More precisely, the mapping $I: H_{\text{dR}}^k(M) \rightarrow H_k(M; \mathbb{R})^*$ is defined as follows: if $[\omega] \in H_{\text{dR}}^k(M)$ is represented by the closed form $\omega \in \Omega^k(M)$, then $I([\omega]) \in H_k(M; \mathbb{R})^*$ is given by

$$I([\omega])([C]) := \int_C \omega$$

for any $[C] \in H_k(M; \mathbb{R})$ which is represented by the cycle $C \subset M$. It's (hopefully!) clear that $I([\omega])$ is a linear functional (provided the integral here is anything sensible), but probably less obvious that I is well-defined and an isomorphism. Of course, to even make sense of the definition of I , we need to talk about integration.

2.9 Integration on Manifolds

As suggested in [Section 2.1](#), the point of defining differential forms as alternating, multilinear gadgets is because they transform in the right way.

2.9.1 Digression on Differentials, Pullbacks, and Dual Maps

First, a quick refresher on *dual maps* or *algebraic adjoints*:

Given vector spaces U and V and a linear map $h: U \rightarrow V$, there is an associated dual map $h^*: V^* \rightarrow U^*$ defined as follows: if $\alpha \in V^*$, then $h^*(\alpha)$ is supposed to be an element of U^* ; that is, a linear functional on U , meaning that we can specify it by specifying what it does to an arbitrary element $u \in U$:

$$h^*(\alpha)(u) := \alpha(h(u)).$$

In other words, $h^*\alpha = \alpha \circ h$. It's easy to check that h^* is linear since h is.

Moreover, when U and V are finite-dimensional with chosen bases u_1, \dots, u_m for U and v_1, \dots, v_n for V , then the matrix for h^* with respect to the dual bases u_1^*, \dots, u_m^* for U^* and v_1^*, \dots, v_n^* for V^* is equal to the transpose (or conjugate transpose, for complex vector spaces) of the matrix for h with respect to the given bases for U and V . This explains the “algebraic adjoint” terminology, since the transpose (or conjugate transpose) is the usual adjoint, at least when the chosen bases are orthonormal. (And I claim it is occasionally useful to think about transpose in these terms, even though in applications we often think about the transpose as just a map $V \rightarrow U$.)

Now, consider a smooth map $g: M \rightarrow N$ and let $p \in M$. Then the differential $dg_p: T_p M \rightarrow T_{g(p)} N$ must have an associated dual map $(dg_p)^*: (T_{g(p)} N)^* \rightarrow (T_p M)^*$.

First of all, notice that $(T_{g(p)} N)^* = \wedge^1((T_{g(p)} N)^*)$ and $(T_p M)^* = \wedge^1((T_p M)^*)$, so the pullback (at p)

$$g_p^*: \wedge^1((T_{g(p)} N)^*) \rightarrow \wedge^1((T_p M)^*)$$

has the same domain and range as $(dg_p)^*$.

Moreover, by definition of the dual map, for a linear functional $\alpha \in (T_{g(p)} N)^*$ and a tangent vector $v \in T_p M$,

$$(dg_p)^*(\alpha)(v) = \alpha(dg_p(v)).$$

But this is exactly how we defined the pullback, where we now think of $\alpha = \omega_{g(p)}$, the value of the 1-form ω at $g(p)$.

In other words, g^* as defined in [Section 2.6](#) is the same as the dual map $(dg)^*$ associated to the differential dg .

2.9.2 How Differential Forms Transform

The goal is to see that transformations of differential forms under changes of coordinates are just like transformations of integrands under changes of coordinates.

As in [Section 2.1](#), let $g: B \rightarrow A$ be a diffeomorphism of open subsets of \mathbb{R}^n and let $f: A \rightarrow \mathbb{R}$ be a smooth function. Define

$$\omega = f dx_1 \wedge \cdots \wedge dx_n \in \Omega^n(A),$$

which pulls back to

$$g^*\omega = (f \circ g)g^*(dx_1) \wedge \cdots \wedge g^*(dx_n) = (f \circ g)(dg)^*(dx_1) \wedge \cdots \wedge (dg)^*(dx_n),$$

where here $(dg)^*$ means the dual of the differential dg . By Homework 2, Problem 1(c), the above is equal to

$$(f \circ g)(\det(dg)^*)dy_1 \wedge \cdots \wedge dy_n,$$

where we recall that y_1, \dots, y_n were just the names of the coordinates on B . Since determinants are invariant under taking transposes, $\det(dg)^* = \det dg = \det Jg$, so this parallels the usual vector calculus change of variables formula, at least when f is smooth.

In other words, differential forms transform in the right way to serve as integrands.

Example 2.9.1. Let $M = (0, +\infty) \times (-\pi, \pi)$, $N = \mathbb{R}^2$, and the map $g: M \rightarrow N$ given by

$$g(r, \theta) := (r \cos \theta, r \sin \theta).$$

In other words, this is the change of variables map between polar coordinates (r, θ) and Cartesian coordinates (x, y) on the plane.

Now, consider the standard area form $\omega = dx \wedge dy$ on N and we want to work out $g^*\omega$ from the definition.

By definition of the differential (Definition 1.3.1), we have that, for $p = (\theta_0, r_0) \in M$,

$$\begin{aligned} dg_p \left(\frac{\partial}{\partial \theta} \right) &= \frac{d}{dt} \Big|_{t=0} g(\theta_0 + t, r_0) = \frac{d}{dt} \Big|_{t=0} r_0(\cos(\theta_0 + t), \sin(\theta_0 + t)) \\ &= r_0(-\sin \theta_0, \cos \theta_0) = r_0 \left(-\sin \theta_0 \frac{\partial}{\partial x} + \cos \theta_0 \frac{\partial}{\partial y} \right) \end{aligned}$$

where I move back and forth between tuple representations and local coordinate basis representations. Similarly,

$$\begin{aligned} dg_p \left(\frac{\partial}{\partial r} \right) &= \frac{d}{dt} \Big|_{t=0} g(\theta_0, r_0 + t) = \frac{d}{dt} \Big|_{t=0} (r_0 + t)(\cos \theta_0, \sin \theta_0) \\ &= (\cos \theta_0, \sin \theta_0) = \cos \theta_0 \frac{\partial}{\partial x} + \sin \theta_0 \frac{\partial}{\partial y}. \end{aligned}$$

Therefore,

$$\begin{aligned} (g^*\omega)_p \left(a \frac{\partial}{\partial \theta} + b \frac{\partial}{\partial r}, c \frac{\partial}{\partial \theta} + d \frac{\partial}{\partial r} \right) &= \omega_{g(p)} \left(dg_p \left(a \frac{\partial}{\partial \theta} + b \frac{\partial}{\partial r} \right), dg_p \left(c \frac{\partial}{\partial \theta} + d \frac{\partial}{\partial r} \right) \right) \\ &= dx \wedge dy \left((b \cos \theta_0 - ar_0 \sin \theta_0) \frac{\partial}{\partial x} + (b \sin \theta_0 + ar_0 \cos \theta_0) \frac{\partial}{\partial y}, \right. \\ &\quad \left. (d \cos \theta_0 - cr_0 \sin \theta_0) \frac{\partial}{\partial x} + (d \sin \theta_0 + cr_0 \cos \theta_0) \frac{\partial}{\partial y} \right) \\ &= (b \cos \theta_0 - ar_0 \sin \theta_0)(d \sin \theta_0 + cr_0 \cos \theta_0) \\ &\quad - (b \sin \theta_0 + ar_0 \cos \theta_0)(d \cos \theta_0 - cr_0 \sin \theta_0) \\ &= r(bc - ad). \end{aligned}$$

On the other hand,

$$dr \wedge d\theta \left(a \frac{\partial}{\partial \theta} + b \frac{\partial}{\partial r}, c \frac{\partial}{\partial \theta} + d \frac{\partial}{\partial r} \right) = bc - ad,$$

so we can conclude that $g^*\omega = r dr \wedge d\theta$.

Since the composition of the constant function 1 (on N) with g is just the constant function 1 (on M), the above discussion implies that

$$g^*\omega = g^*(dx \wedge dy) = (\det Jg)dr \wedge d\theta,$$

where $\det Jg$ is the Jacobian determinant

$$\begin{vmatrix} \frac{\partial g_1}{\partial r} & \frac{\partial g_1}{\partial \theta} \\ \frac{\partial g_2}{\partial r} & \frac{\partial g_2}{\partial \theta} \end{vmatrix} = \begin{vmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{vmatrix} = r \cos^2 \theta + r \sin^2 \theta = r.$$

So the above discussion of how differential forms transform matches up to the standard vector calculus fact that the Cartesian area element $dx dy$ transforms to the polar area element $r dr d\theta$.

2.9.3 Definition of the Integral

First, we define integration of forms on open sets in \mathbb{R}^n : if $A \subset \mathbb{R}^n$ is open, then any $\omega \in \Omega^n(A)$ can be written as

$$\omega = f dx_1 \wedge \cdots \wedge dx_n,$$

where f is a smooth function on A . Define

$$\int_A \omega = \int_A f dx_1 \cdots dx_n,$$

where the integral on the right is just the usual Riemann integral of the (smooth) function f over A .

By the previous discussion, if $g: B \rightarrow A$ is an orientation-preserving diffeomorphism, then $\det Jg > 0$, so

$$\begin{aligned} \int_A \omega &= \int_A f dx_1 \cdots dx_n = \int_B (f \circ g) |\det Jg| dy_1 \cdots dy_n = \int_B (f \circ g)(\det Jg) dy_1 \cdots dy_n \\ &= \int_B (f \circ g)(\det Jg) dy_1 \wedge \cdots \wedge dy_n = \int_B g^* \omega. \end{aligned}$$

In other words, our definition is actually coordinate-independent, as you would hope.

If g is orientation-reversing, then we get

$$\int_A \omega = - \int_B g^* \omega.$$

Since we have no standard coordinates on an arbitrary manifold, these transformations are essential to defining integrals on manifolds.

Suppose next that we have an n -dimensional manifold M and the support of $\omega \in \Omega^n(M)$ is contained in a single coordinate chart (U, ϕ) so that $\phi: U \rightarrow \phi(U)$ is orientation-preserving. Then $\phi^* \omega \in \Omega^n(U)$, which we know how to integrate, so define

$$\int_M \omega = \int_U \phi^* \omega.$$

Proposition 2.9.2. *The above is well-defined.*

Proof. We need to check that, if the support of ω is also contained in some other coordinate chart (V, ψ) so that $\psi: V \rightarrow \psi(V)$ is orientation-preserving, then we get the same answer whether we compute $\int_M \omega = \int_U \phi^* \omega$ or $\int_M \omega = \int_V \psi^* \omega$.

We can move between U and V by way of the map $\psi^{-1} \circ \phi: U \rightarrow V$, which is orientation-preserving, so

$$\int_V \psi^* \omega = \int_U (\psi^{-1} \circ \phi)^* \psi^* \omega = \int_U \phi^*(\psi^{-1})^* \psi^* \omega = \int_U \phi^*(\psi \circ \psi^{-1}) * \omega = \int_U \phi^* \omega,$$

where I'm using the previously unstated fact that, for maps $f: A \rightarrow B$ and $g: B \rightarrow C$, $(g \circ f)^* = f^* g^*$, which follows from the chain rule. \square

So we've now succeeded in defining the integral of a form contained entirely in a single local coordinate chart. But what about arbitrary forms?

First of all, to avoid worrying about convergence issues, let's restrict to *compactly supported* forms: that is, those forms whose support (the set of points on which they are non-vanishing) is contained in some compact subset of M . Of course, if M is itself compact, this is no restriction at all.

So now suppose we have a compactly supported form $\omega \in \Omega^n(M)$. To define $\int_M \omega$, we will use a partition of unity argument.

Definition 2.9.3. Let T be a topological space and let $\{W_\alpha\}_{\alpha \in \mathcal{I}}$ be an open cover of T , where \mathcal{I} is some indexing set. A *partition of unity* subordinate to $\{W_\alpha\}_{\alpha \in \mathcal{I}}$ is a collection $\{\rho_\alpha\}_{\alpha \in \mathcal{I}}$ of continuous functions $\rho_\alpha: T \rightarrow [0, 1]$ so that $\text{supp}(\rho_\alpha) \subset W_\alpha$ (i.e., $\rho_\alpha(x) = 0$ for all $x \notin W_\alpha$) and, for all $x \in T$,

- there exists a neighborhood of x on which all but finitely many of the ρ_α are identically zero and
- $\sum_{\alpha \in \mathcal{I}} \rho_\alpha(x) = 1$ (note that the previous condition ensures this sum is finite and hence converges).

These always exist in reasonable topological spaces:

Theorem 2.9.4. *Every open cover on a paracompact Hausdorff space has a subordinate partition of unity.*

Since manifolds are second countable and Hausdorff, they are paracompact, so this implies (after modifying the proof to make the ρ_α smooth):

Corollary 2.9.5. *Every manifold has a partition of unity subordinate to its coordinate atlas so that the functions are smooth.*

Now, finally, we can define integration of compactly-supported forms on manifolds:

Definition 2.9.6. Let M be an n -dimensional manifold and let $\omega \in \Omega^n(M)$ be compactly supported. Let $\{\rho_\alpha\}$ be a partition of unity subordinate to the (maximal) atlas $\{(U_\alpha, \phi_\alpha)\}$ on M and define

$$\int_M \omega = \sum_{\alpha} \int_M \rho_\alpha \omega.$$

Remark 2.9.7. (i) $\rho_\alpha \omega$ is only non-vanishing inside the support of ρ_α , which is contained in some coordinate neighborhood $\phi_\alpha(U_\alpha)$, so the integrals inside the sum are well-defined.

- (ii) The local finiteness of the partition of unity guarantees only finitely many terms are nonzero on the compact support of ω , so the sum converges.
- (iii) If ω is already contained in some coordinate chart (ϕ, U) , then this definition agrees with the previous one: since $\sum_{\alpha} \rho_{\alpha}(p) = 1$ for each $p \in M$, we have that $\sum_{\alpha} \rho_{\alpha}\omega = \omega$, so using the previous definition

$$\int_M \omega = \int_U \phi^* \omega = \int_U \phi^*(\sum_{\alpha} \rho_{\alpha}\omega) = \sum_{\alpha} \int_U \phi^*(\rho_{\alpha}\omega) = \sum_{\alpha} \int_M \rho_{\alpha}\omega$$

by linearity of pullback and of integration on \mathbb{R}^n .

- (iv) If $\{\sigma_{\alpha}\}$ were a different subordinate partition of unity, then the previous observation implies that, or each α ,

$$\int_M \rho_{\alpha}\omega = \sum_{\beta} \int_M \sigma_{\beta}\rho_{\alpha}\omega$$

and similarly for each β

$$\int_M \sigma_{\beta}\omega = \sum_{\alpha} \int_M \rho_{\alpha}\sigma_{\beta}\omega,$$

so

$$\sum_{\alpha} \int_M \rho_{\alpha}\omega = \sum_{\alpha} \sum_{\beta} \int_M \sigma_{\beta}\rho_{\alpha}\omega = \sum_{\beta} \sum_{\alpha} \int_M \rho_{\alpha}\sigma_{\beta}\omega = \sum_{\beta} \int_M \sigma_{\beta}\omega,$$

which means we get the same value for $\int_M \omega$ no matter which partition of unity we use.

- (v) Integration is a linear operator: for $a \in \mathbb{R}$ and $\omega, \eta \in \Omega^n(M)$, $\int_M (\omega + a\eta) = \int_M \omega + a \int_M \eta$.
- (vi) If $\omega \in \Omega^n(M)$ and $g : N \rightarrow M$ is an orientation-preserving diffeomorphism, then

$$\int_M \omega = \int_N g^* \omega.$$

Example 2.9.8. Let $\gamma : [a, b] \rightarrow M$ be a smooth curve with $\gamma(a) = p$ and $\gamma(b) = q$. Assume $f \in C^{\infty}(M)$ and $\omega = df \in \Omega^1(M)$. Then I claim that

$$\int_{[a,b]} \gamma^* \omega = f(q) - f(p).$$

This is just a computation:

$$\begin{aligned} \int_{[a,b]} \gamma^* \omega &= \int_{[a,b]} \gamma^*(df) = \int_{[a,b]} d(\gamma^* f) = \int_{[a,b]} d(f \circ \gamma) = \int_{[a,b]} (f \circ \gamma)'(t) dt \\ &= (f \circ \gamma)(b) - (f \circ \gamma)(a) = f(q) - f(p) \end{aligned}$$

where the second equality follows from Proposition 2.6.4, the third from Example 2.6.2, and the fifth from the fundamental theorem of calculus.

Example 2.9.9. Suppose $\gamma: S^1 \rightarrow M$ is smooth and $\omega \in \Omega^1(M)$. Then we define the *line integral* of ω around γ to be

$$\oint_{\gamma} \omega := \int_{S^1} \gamma^* \omega.$$

Now suppose $M = \mathbb{R}^n$, so that $\omega = a_1 dx_1 + \cdots + a_n dx_n$. Letting $\{\frac{\partial}{\partial \theta}\}$ be the local coordinate basis for the tangent space $T_t S^1$ at a point $t \in S^1$ and writing $\gamma(t) = (\gamma_1(t), \dots, \gamma_n(t))$, we can compute

$$(\gamma^* dx_i)_t \left(b \frac{\partial}{\partial \theta} \right) = (dx_i)_{\gamma(t)} \left(d\gamma_t \left(b \frac{\partial}{\partial \theta} \right) \right) = b(dx_i)_{\gamma(t)} \left(d\gamma_t \left(\frac{\partial}{\partial \theta} \right) \right) = b(dx_i)_{\gamma(t)} \left(\frac{\partial}{\partial \theta}(\gamma)(t) \right)$$

using Lemma 1.3.4. But then

$$\frac{\partial}{\partial \theta}(\gamma)(t) = \gamma'_1(t) \frac{\partial}{\partial x_1} + \cdots + \gamma'_n(t) \frac{\partial}{\partial x_n},$$

so we see that

$$(\gamma^* dx_i)_t \left(b \frac{\partial}{\partial \theta} \right) = b\gamma'_i(t)$$

or, in other words, $(\gamma^* dx_i)_t = \gamma'_i(\theta)d\theta$. Therefore,

$$(\gamma^* \omega)_t = (\gamma^* a_1)_t (\gamma^* dx_1)_t + \cdots + (\gamma^* a_n)_t (\gamma^* dx_n)_t = [a_1(\gamma(t))\gamma'_1(t) + \cdots + a_n(\gamma(t))\gamma'_n(t)] d\theta.$$

Notice that the coefficient is just $A(\gamma(t)) \cdot \gamma'(t)$, where the vector field $A = (a_1, \dots, a_n)$ is dual to ω . Of course, this is the sort of expression you usually see in the Multivariable Calculus definition of a line integral.

Putting Examples 2.9.8 and 2.9.9 together, notice that if $\omega = df \in \Omega^1(M)$, then $\oint_{\gamma} \omega = 0$ for all closed curves γ in M .

Proposition 2.9.10. *The converse is true. That is, any $\omega \in \Omega^1(M)$ with the property that $\oint_{\gamma} \omega = 0$ for all closed curves γ in M is exact.*

Exercise 2.9.11. Prove this.

2.10 Integration on Submanifolds

So far, we've only defined integrating an n -form over an n -dimensional manifold. However, notice that any open subset of an n -dimensional manifold is also an n -manifold, so we can actually integrate over arbitrary open subsets.

Indeed, given an n -manifold M and any $\omega \in \Omega^n(M)$, we get a (signed) measure m on open subsets of M defined by

$$m(A) := \int_A \omega.$$

When ω is a volume form (meaning it is nowhere-vanishing), then $m(A)$ will have the same sign for all open A , and if the sign is positive then m will satisfy the axioms of a measure on the σ -algebra of Borel sets on M ; that is, it is a Borel measure. Moreover, if M is compact then $\int_M \omega$ is finite, so m can be normalized to

give a probability measure. Indeed, starting with a volume form is by far the most common way of defining probability measures on (orientable) manifolds.

However, this is not so useful for the integrals mentioned in [Section 2.8](#) that show up in de Rham cohomology. After all, if N is a k -dimensional submanifold of the n -dimensional manifold M with $k < n$ and $\omega \in \Omega^n(M)$, then $\int_N \omega = 0$.¹²

Rather, for de Rham cohomology purposes, we need a way of defining $\int_N \eta$ where $\eta \in \Omega^k(M)$. Intuitively, we want to “restrict” η to N here, but trying to write this intuition out directly leads to something quite complicated. But this is a place where all our effort in defining things in such a relatively abstract way pays off by giving us a simple and natural (though seemingly slightly indirect) way of defining our integral:

Definition 2.10.1. Suppose N is a k -dimensional submanifold of the manifold M and let $i : N \hookrightarrow M$ be the inclusion map. If $\eta \in \Omega^k(M)$, define

$$\int_N \eta := \int_N i^* \eta.$$

The point here is that $i^* \eta \in \Omega^k(N)$, and [Definition 2.9.6](#) applies.

Example 2.10.2. Let $M = \mathbb{R}^3$ and let S be a surface which is the graph of a function; i.e.,

$$S = \{(u, v, g(u, v)) : u, v \in \mathbb{R}\},$$

for some smooth function $g : \mathbb{R}^2 \rightarrow \mathbb{R}$ (this is not really a restrictive assumption: the implicit function theorem implies that locally every surface looks like this, possibly after permuting coordinates). We can parametrize S by the map $h : \mathbb{R}^2 \rightarrow \mathbb{R}^3$ given by $h(u, v) = (u, v, g(u, v))$.

Suppose $\omega \in \Omega^2(\mathbb{R}^3)$ is compactly supported. We can write

$$\omega = f_1 dy \wedge dz + f_2 dz \wedge dx + f_3 dx \wedge dy.$$

Then

$$\int_S \omega = \int_{\mathbb{R}^2} h^* \omega,$$

so we can compute

$$\begin{aligned} h^* dx &= d(x \circ h) = du \\ h^* dy &= d(y \circ h) = dv \\ h^* dz &= d(z \circ h) = d(g(u, v)) = \frac{\partial g}{\partial u} du + \frac{\partial g}{\partial v} dv, \end{aligned}$$

and hence

$$\begin{aligned} h^*(dx \wedge dy) &= h^* dx \wedge h^* dy = du \wedge dv \\ h^*(dy \wedge dz) &= h^* dy \wedge h^* dz = dv \wedge \left(\frac{\partial g}{\partial u} du + \frac{\partial g}{\partial v} dv \right) = -\frac{\partial g}{\partial u} du \wedge dv \\ h^*(dz \wedge dx) &= h^* dz \wedge h^* dx = \left(\frac{\partial g}{\partial u} du + \frac{\partial g}{\partial v} dv \right) \wedge du = -\frac{\partial g}{\partial v} du \wedge dv. \end{aligned}$$

¹²As written, this doesn't quite make sense given our definitions, but one way to formalize it would be to realize N as the intersection of a decreasing nested sequence of open sets and then taking the limit of the integral of ω over those open sets. The measures of these sets will go to zero.

Therefore,

$$\int_S \omega = \int_{\mathbb{R}^2} h^* \omega = \int_{\mathbb{R}^2} \left(-\frac{\partial g}{\partial u} f_1 - \frac{\partial g}{\partial v} f_2 + f_3 \right) du \wedge dv. \quad (2.6)$$

Now, I claim this is a standard formula from vector calculus. To see this, let $\vec{F} = (f_1, f_2, f_3)$ be the vector field corresponding to ω (in the notation from [Section 2.2](#), $\omega = \star F^\flat$) and let $\vec{n} = \left(-\frac{\partial g}{\partial u}, -\frac{\partial g}{\partial v}, 1 \right)$. Notice that, at $p = (u, v, g(u, v)) \in S$, the tangent space $T_p S$ is spanned by

$$\frac{d}{dt} \Big|_{t=0} h(t, 0) = \frac{d}{dt} \Big|_{t=0} (t, 0, g(t, 0)) = \left(1, 0, \frac{\partial g}{\partial u} \right)$$

and

$$\frac{d}{dt} \Big|_{t=0} h(0, t) = \frac{d}{dt} \Big|_{t=0} (0, t, g(0, t)) = \left(0, 1, \frac{\partial g}{\partial v} \right).$$

Since

$$\vec{n} \cdot \left(1, 0, \frac{\partial g}{\partial u} \right) = \left(-\frac{\partial g}{\partial u}, -\frac{\partial g}{\partial v}, 1 \right) \cdot \left(1, 0, \frac{\partial g}{\partial u} \right) = -\frac{\partial g}{\partial u} + \frac{\partial g}{\partial u} = 0$$

and

$$\vec{n} \cdot \left(0, 1, \frac{\partial g}{\partial v} \right) = \left(-\frac{\partial g}{\partial u}, -\frac{\partial g}{\partial v}, 1 \right) \cdot \left(0, 1, \frac{\partial g}{\partial v} \right) = -\frac{\partial g}{\partial v} + \frac{\partial g}{\partial v} = 0,$$

the vector \vec{n} is normal to S at all points of S .

Therefore, if $\vec{u} = \frac{\vec{n}}{\|\vec{n}\|}$ and $dA = \|\vec{n}\| du \wedge dv$ (usually called the area form or area element for S), then our formula (2.6) can be re-written as

$$\int_S \omega = \int_{\mathbb{R}^2} (\vec{F} \cdot \vec{u}) dA,$$

which is the standard formula from vector calculus for computing the surface integral of a vector field.

Example 2.10.3. Recall the form $\omega = \frac{x dy - y dx}{x^2 + y^2} \in \Omega^1(\mathbb{R}^2 - \{0\})$ defined in [Exercise 2.5.8](#) and also considered in [Example 2.8.2](#). Let $f: S^1 \rightarrow \mathbb{R}^2 - \{0\}$ defined by $f(\theta) = (\cos \theta, \sin \theta)$ be the standard parametrization of the unit circle. Then

$$\int_{S^1} \omega = \int_{S^1} f^* \omega.$$

We saw in [Example 2.6.3](#) that $f^* dx = -\sin \theta d\theta$ and a similar computation shows that $f^* dy = \cos \theta d\theta$. Since

$$\omega_{f(\theta_0)} = \frac{\cos(\theta_0) dy - \sin(\theta_0) dx}{\cos^2 \theta_0 + \sin^2 \theta_0} = \cos(\theta_0) dy - \sin(\theta_0) dx,$$

we see that

$$(f^* \omega)_{\theta_0} = (\cos^2 \theta_0 + \sin^2 \theta_0) d\theta = d\theta,$$

so

$$\int_{S^1} \omega = \int_{S^1} f^* \omega = \int_{S^1} d\theta = 2\pi.$$

2.11 Stokes' Theorem

At this point we have all the ingredients in place to prove (a version of) de Rham's theorem: for a k -dimensional homology class $[C] \in H_k(M; \mathbb{R})$ represented by a submanifold C , and a de Rham cohomology class $[\omega]$ represented by a closed form $\omega \in \Omega^k(M)$, we at least know what

$$I([\omega])([C]) := \int_C \omega$$

means.

However, to see that this is well-defined, we need to know that, if $[\omega_1] = [\omega_2]$, then $\int_C \omega_1 = \int_C \omega_2$, and similarly that if $[C_1] = [C_2]$, then $\int_{C_1} \omega = \int_{C_2} \omega$.

For the first, suppose $\omega_1, \omega_2 \in \Omega^k(M)$ represent the same de Rham cohomology class (i.e., the same element of $H_{\text{dR}}^k(M) = \frac{\ker d}{\text{im } d}$), then $\omega_1 - \omega_2 = d\eta$ for some $\eta \in \Omega^{k-1}(M)$. So we need to show that

$$0 = \int_C d\eta = \int_C (\omega_1 - \omega_2) = \int_C \omega_1 - \int_C \omega_2.$$

Since C is a cycle, $\partial C = \emptyset$, so this will follow from:

Theorem 2.11.1 (Stokes' Theorem for Differential Forms). *If M is a compact, oriented n -manifold with boundary and $\omega \in \Omega^{n-1}(M)$, then*

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

As for the second equality we need to show that the pairing I is well defined, suppose $[C_1] = [C_2]$. This means¹³ that $\partial E = C_2 - C_1$ for some $(k+1)$ -chain E . But then for any closed $\omega \in \Omega^k(M)$,

$$\int_{C_2} \omega - \int_{C_1} \omega = \int_{C_2 - C_1} \omega = \int_E d\omega = \int_E 0 = 0$$

by [Theorem 2.11.1](#).

In other words, Stokes' Theorem is the key to showing that the pairing I is well-defined. Of course, to really properly prove [Theorem 2.8.3](#) we would also need to argue that every homology class can be represented by a submanifold (or at least a finite union of submanifolds), which we're not going to bother with.

To give a complete proof of [Theorem 2.11.1](#), we would need to develop the theory of manifolds with boundary. However, since we won't really be using that theory for anything else, I don't think it's worth the effort. Instead, I'll just give the part of the proof that doesn't interact with the boundary. See Guillemin and Pollack [4] for details on manifolds with boundary and a complete proof of Stokes' Theorem.

First part of proof of Theorem 2.11.1. Notice, first of all, that by linearity of the integral we can assume ω has compact support contained in the image of some local coordinate chart (ϕ, U) , so that $\int_M d\omega = \int_U \phi^* d\omega = \int_U d\phi^* \omega$. If we assume $\phi(U) \cap \partial M = \emptyset$, then $\int_{\partial M} \omega = 0$, so the goal is to show that $\int_U d\phi^* \omega = 0$.

We know that $\phi^* \omega \in \Omega^{n-1}(U) \subset \Omega^{n-1}(\mathbb{R}^n)$, so we can write

$$\phi^* \omega = \sum_{i=1}^n (-1)^{i-1} f_i dx_1 \wedge \cdots \wedge d\check{x}_i \wedge \cdots \wedge dx_n$$

¹³If you haven't seen homology before, you might just need to trust me on this.

for some compactly supported functions f_i , where $d\check{x}_i$ means that dx_i is omitted and the signs are chosen so that

$$d\phi^*\omega = \left(\sum_{i=1}^n \frac{\partial f_i}{\partial x_i} \right) dx_1 \wedge \cdots \wedge dx_n.$$

Since the f_i are compactly supported, we can extend their domain of definition to all of \mathbb{R}^n by defining them to be zero outside U . Therefore,

$$\int_U d\phi^*\omega = \int_U \left(\sum_{i=1}^n \frac{\partial f_i}{\partial x_i} \right) dx_1 \wedge \cdots \wedge dx_n = \int_{\mathbb{R}^n} \left(\sum_{i=1}^n \frac{\partial f_i}{\partial x_i} \right) dx_1 \wedge \cdots \wedge dx_n = \sum_{i=1}^n \int_{\mathbb{R}^n} \frac{\partial f_i}{\partial x_i} dx_1 \cdots dx_n.$$

By Fubini's Theorem, we can interpret each term as an iterated integral and do the integration in any order we like:

$$\int_{\mathbb{R}^n} \frac{\partial f_i}{\partial x_i} dx_1 \cdots dx_n = \int_{\mathbb{R}^{n-1}} \left(\int_{-\infty}^{\infty} \frac{\partial f_i}{\partial x_i} dx_i \right) dx_1 \cdots d\check{x}_i \cdots dx_n.$$

Now we're thinking of the integrand $\int_{-\infty}^{\infty} \frac{\partial f_i}{\partial x_i} dx_i$ as a function of $(x_1, \dots, \check{x}_i, \dots, x_n)$. Which function? Well, to any specific point $(a_1, \dots, \check{a}_i, \dots, a_n)$ it assigns the number $\int_{-\infty}^{\infty} g'(t) dt$ where $g(t) = f_i(a_1, \dots, t, \dots, a_n)$. But since f_i is compactly supported, we know there exists T big enough that $g(t) = 0$ for all $t \geq T$.

Then the Fundamental Theorem of Calculus implies that

$$\int_{-\infty}^{\infty} g'(t) dt = \int_{-T}^T g'(t) dt = g(T) - g(-T) = 0 - 0 = 0,$$

so we have

$$\int_M d\omega = \int_U \phi^* d\omega = \int_U d\phi^*\omega = 0$$

as desired. \square

To give the rest of the proof of [Theorem 2.11.1](#), one needs to deal with the case $\phi(U) \cap \partial M \neq \emptyset$. Again, see Guillemin and Pollack [4] if you're interested.

Example 2.11.2. Suppose S is a smooth surface in \mathbb{R}^3 with connected boundary $\partial S = \Gamma$, which is parametrized by $\gamma: S^1 \rightarrow \mathbb{R}^3$. Let $\omega \in \Omega^1(\mathbb{R}^3)$. Then [Theorem 2.11.1](#) tells us that

$$\int_S d\omega = \int_{\Gamma} \omega. \tag{2.7}$$

Let's translate this into vector calculus language. Let $\vec{F} \in \mathfrak{X}(\mathbb{R}^3)$ be the vector field corresponding to ω . That is, if $\omega = f dx + g dy + h dz$, then $\vec{F} = f \frac{\partial}{\partial x} + g \frac{\partial}{\partial y} + h \frac{\partial}{\partial z}$.

Following [Example 2.9.9](#),

$$\int_{\Gamma} \omega = \int_{S^1} \gamma^* \omega = \int_{S^1} \vec{F} \cdot \gamma'(t) dt$$

which we often write as $\oint_{\Gamma} \vec{F} \cdot d\Gamma$.

To translate the left-hand side of (2.7), recall that we saw in Section 2.2 that the vector field corresponding to $d\omega$ is $\nabla \times \vec{F}$. Then, following Example 2.10.2 we have

$$\int_S d\omega = \int_{\mathbb{R}^2} ((\nabla \times \vec{F}) \cdot u) dA,$$

which is what we mean in vector calculus by the notation $\iint_S (\nabla \times \vec{F}) \cdot dS$, where dS is the area element on S .

In other words, translating the Stokes' theorem result (2.7) into vector calculus language recovers the vector calculus Stokes' Theorem (or Curl Theorem)

$$\iint_S (\nabla \times \vec{F}) \cdot dS = \oint_{\Gamma} \vec{F} \cdot d\Gamma.$$

Exercise 2.11.3. Do the same sort of translation for the Divergence Theorem.

Chapter 3

Lie Groups and Homogeneous Spaces

3.1 Some Motivation from de Rham Cohomology

Recall that we have the de Rham cohomology group $H_{\text{dR}}^k(M)$, which tells us a lot about the topology of the manifold, but can in general be quite challenging to compute. After all, the $\Omega^k(M)$ are $C^\infty(M)$ -modules, and hence infinite-dimensional as real vector spaces. If M is, e.g., compact then $H_{\text{dR}}^k(M)$ will be a finite-dimensional real vector space, but it appears as the quotient of two infinite-dimensional vector spaces, so it can be somewhat challenging to compute.

However, when the manifold has lots of symmetry, computing the de Rham cohomology groups can actually become a finite-dimensional problem. In this case, a reasonable interpretation of “lots of symmetry” is “admits a transitive action of a Lie group,” resulting in a class of manifolds called *homogeneous spaces*. We haven’t defined Lie groups yet (though we will shortly), but the idea is that it’s a group which is also a manifold in a sensible way. At a more conceptual level, it’s just a continuous family of symmetries of the manifold, where here “continuous” means they form a continuum.

The following statement says more precisely how de Rham cohomology becomes finite-dimensional in this setting, though don’t worry too much about terms in the statement that we haven’t defined yet.

Theorem 3.1.1. *Let M be a homogeneous space where the transitive action is by a compact, connected Lie group. Then:*

- *Each closed k -form on M is in the same de Rham cohomology class as a G -invariant closed k -form.*
- *If a G -invariant closed k -form is exact, then it is also the exterior derivative of some G -invariant $(k-1)$ -form.*

Corollary 3.1.2. *The cohomology of G -invariant forms on M is isomorphic to its de Rham cohomology.*

Since, as we will see, G -invariant forms are determined by their behavior at a single point, this means that the space of G -invariant forms will be finite-dimensional (it can’t be bigger than $\wedge^k(T_p M)^*$), and hence computing de Rham cohomology becomes a finite-dimensional linear algebra problem.

Of course, the simplest possible example of a homogeneous space is a manifold that is itself a group, which acts by itself by (e.g. left) multiplication.

Example 3.1.3. S^3 is a group: it's just the unit quaternions, and hence it has a group operation given by multiplication of quaternions. Moreover, the vector fields $X, Y, Z \in \mathfrak{X}(S^3)$ from [Section 1.8](#) given by

$$X(p) = pi, \quad Y(p) = pj, \quad Z(p) = pk$$

are invariant under left-multiplication, and hence the dual 1-forms $\alpha, \beta, \gamma \in \Omega^1(S^3)$ from [Example 2.7.6](#) are also invariant under left-multiplication. Moreover, it is easy to show that *all* left-invariant 1-forms must be of the form $a\alpha + b\beta + c\gamma$ for constants $a, b, c \in \mathbb{R}$, so the space of left-invariant 1-forms is 3-dimensional.

Likewise, $\alpha \wedge \beta, \beta \wedge \gamma, \gamma \wedge \alpha$ form a basis for the space of left-invariant 2-forms on S^3 , which is therefore also 3-dimensional.

We saw in [Example 2.7.6](#) that

$$d\alpha = -2\beta \wedge \gamma, \quad d\beta = -2\gamma \wedge \alpha, \quad d\gamma = -2\alpha \wedge \beta,$$

so there are *no* closed left-invariant 1-forms, and [Corollary 3.1.2](#) implies $H_{\text{dR}}^1(S^3) = \{0\}$.

On the other hand,

$$d(\beta \wedge \gamma) = d\beta \wedge \gamma - \beta \wedge d\gamma = (-2\gamma \wedge \alpha) \wedge \gamma - \beta \wedge (-2\alpha \wedge \beta) = 0 - 0 = 0$$

and likewise for the other left-invariant 2-forms, so the left-invariant 2-forms are all closed... but they're also all exact: for example, $\beta \wedge \gamma = d\left(-\frac{1}{2}\alpha\right)$. So this tells us that $H_{\text{dR}}^2(S^3) = \{0\}$.

Therefore, the cohomology of left-invariant forms on S^3 is simply

$$\begin{aligned} H_{\text{dR}}^0(S^3) &\cong \mathbb{R} \\ H_{\text{dR}}^1(S^3) &= \{0\} \\ H_{\text{dR}}^2(S^3) &= \{0\} \\ H_{\text{dR}}^3(S^3) &\cong \mathbb{R}, \end{aligned}$$

which agrees with the usual (simplicial/singular/de Rham/sheaf/whatever) cohomology of S^3 .

The point here is that de Rham cohomology simplifies in the presence of symmetry.

Of course, this is a general principle: typically we expect lots of computations to simplify in the presence of symmetry. Lie groups provide the proper notion of (continuous) symmetries of manifolds, so if we want to understand symmetries and use them to simplify problems, we should probably care at least a little bit about Lie groups.

3.2 Lie groups

The basic idea of a Lie group is that it is a group which is also a manifold, and that the group structure and the manifold structure are compatible. In this setting, compatibility is going to mean that, appropriately interpreted, the group operations are smooth maps.

Before actually stating the definition, the examples to have in mind are matrix groups: $\text{GL}_n(\mathbb{R}), \text{GL}_n(\mathbb{C}), \text{GL}_n(\mathbb{H}), \text{SL}_n(\mathbb{R}), \text{SL}_n(\mathbb{C}), \text{O}(n), \text{SO}(n), \text{U}(n), \text{SU}(n), \text{Sp}(n)$, the Heisenberg group $\left\{ \begin{bmatrix} 1 & a & c \\ 0 & 1 & b \\ 0 & 0 & 1 \end{bmatrix} : a, b, c \in \mathbb{R} \right\}$, etc.

For example, $\mathrm{GL}_n(\mathbb{C})$ is an open subset of $\mathrm{Mat}_{n \times n}(\mathbb{C}) \cong \mathbb{C}^{n^2} \cong \mathbb{R}^{2n^2}$, so it is certainly a $2n^2$ -dimensional manifold, and the entries of the product of two matrices can be written as (quadratic) polynomials in the entries of the terms, so this is smooth. Moreover, by the adjugate formula $A^{-1} = \frac{1}{\det A} \mathrm{adj}(A)$, the entries of A^{-1} are rational functions of the entries of A with non-vanishing denominators, so taking the inverse of a matrix is also a smooth map.

Moreover, as we saw in [Example 1.3.12](#), the unitary group $\mathrm{U}(n)$ is the level set over a regular value of a smooth map defined on $\mathrm{Mat}_{n \times n}(\mathbb{C})$, so this is also a manifold, and the same argument shows that the group operation and taking inverses are smooth.

Definition 3.2.1. A *Lie group* G is a group which is also a differentiable manifold so that the map $G \times G \rightarrow G$ given by $(g, h) \mapsto gh^{-1}$ is smooth.

The map that shows up in this definition is a slightly funny one; it's maybe more common to require that the maps

- $G \times G \rightarrow G$ given by $(g, h) \mapsto gh$ and
- $G \rightarrow G$ given by $g \mapsto g^{-1}$

are smooth.

Exercise 3.2.2. Show that the above two maps being smooth is equivalent to the single map $(g, h) \mapsto gh^{-1}$ being smooth.

Example 3.2.3. Consider the collection of *affine maps* $\varphi_{a,b} : \mathbb{R} \rightarrow \mathbb{R}$ of the real line, given by $\varphi_{a,b}(t) = at + b$.

The composition of two affine maps is affine:

$$(\varphi_{c,d} \circ \varphi_{a,b})(t) = \varphi_{c,d}(at + b) = c(at + b) + d = act + bc + d. \quad (3.1)$$

Also, if $a \neq 0$ then $\varphi_{a,b}$ is invertible, with inverse $\varphi_{a,b}^{-1} = \varphi_{1/a, -b/a}$. For example,

$$(\varphi_{1/a, -b/a} \circ \varphi_{a,b})(t) = \varphi_{1/a, -b/a}(at + b) = \frac{1}{a}(at + b) - \frac{b}{a} = t + \frac{b}{a} - \frac{b}{a} = t.$$

Notice that, in (3.1), if the two maps being composed are invertible (i.e., $a \neq 0$ and $c \neq 0$) then their composition is also invertible ($ac \neq 0$). Therefore, the collection of invertible affine maps forms a group $\mathrm{Aff}(\mathbb{R})$. To see that it is a Lie group, we just need to check that

$$(\varphi_{a,b}, \varphi_{c,d}) \mapsto \varphi_{a,b} \circ \varphi_{c,d}^{-1} = \varphi_{a,b} \circ \varphi_{1/c, -d/c} = \varphi_{a/c, -ad/c+b}$$

is smooth, which it is since a/c and $-ad/c + b$ are smooth functions of a, b, c, d .

Thinking projectively, we can identify \mathbb{R} with the affine subspace $\left\{ \begin{bmatrix} t \\ 1 \end{bmatrix} : t \in \mathbb{R} \right\} \subset \mathbb{R}^2$ and, since

$$\begin{bmatrix} a & b \\ 0 & 1 \end{bmatrix} \begin{bmatrix} t \\ 1 \end{bmatrix} = \begin{bmatrix} at + b \\ 1 \end{bmatrix},$$

we can represent $\mathrm{Aff}(\mathbb{R})$ as the subgroup of $\mathrm{GL}_2(\mathbb{R})$ given by $\left\{ \begin{bmatrix} a & b \\ 0 & 1 \end{bmatrix} : a \neq 0 \right\}$. We can also see that $\mathrm{Aff}(\mathbb{R})$ is a Lie group from this perspective since it is a closed subgroup of the Lie group $\mathrm{GL}_2(\mathbb{R})$.

Definition 3.2.4. Define *left* and *right translation* by $g \in G$ to be the maps $L_g : G \rightarrow G$ and $R_g : G \rightarrow G$ defined by $L_g(h) = gh$ and $R_g(h) = hg$.

Notice that smoothness of L_g and R_g follows from smoothness of $(g, h) \mapsto gh^{-1}$.

3.3 Lie Algebras

One of the beautiful things about Lie groups is that the algebraic structure of the group and the geometric structure of the underlying manifold are closely related, so we can often learn things about the geometry by studying the algebra, or *vice versa*.

For example, the left multiplication map $L_g : G \rightarrow G$ is a diffeomorphism: we already know it's smooth, and its inverse is clearly $L_{g^{-1}}$, which is also smooth. Moreover, if $e \in G$ is the identity element, then $L_g(e) = g$, so the family of L_g diffeomorphisms will map the identity element of G to any other element of G that we like (equivalently, the action of G on itself by left multiplication is transitive). So even though from the algebraic perspective G has a distinguished point e , from the geometric perspective all points look like all other points (since they all look like the identity).

Importantly, $(dL_g)_e : T_e G \rightarrow T_g G$ is a linear isomorphism, so we can explicitly identify every tangent space with the tangent space to the identity. Note, first of all, that this implies the tangent bundle is trivial: $TG \cong G \times T_e G \cong G \times \mathbb{R}^n$, where n is the dimension of the group. Moreover, it tells us that it might be important to understand the tangent space to the identity.

Indeed, as we will shortly see, the tangent space to the identity has extra algebraic structure, and simply-connected Lie groups are completely determined by this algebraic structure.

The specific algebraic structure is that of a *Lie algebra*, which we've encountered before in [Section 1.6](#) when talking about Lie brackets of vector fields, but now we give the formal definition:

Definition 3.3.1. A real vector space \mathfrak{g} is a *Lie algebra* if it has a bilinear operator $[\cdot, \cdot] : \mathfrak{g} \times \mathfrak{g} \rightarrow \mathfrak{g}$ (the bracket) such that, for all $x, y, z \in \mathfrak{g}$,

- $[x, x] = 0$ (alternativity)
- $[[x, y], z] + [[y, z], x] + [[z, x], y] = 0$ (Jacobi identity).

Example 3.3.2. As we saw in [Example 1.6.7](#), \mathbb{R}^3 forms a Lie algebra with the bracket operation $[v, w] = v \times w$.

Example 3.3.3. As we saw in [Example 1.6.10](#), the space $\mathfrak{so}(3)$ of skew-symmetric 3×3 matrices is a Lie algebra with bracket $[A, B] = AB - BA$.

Example 3.3.4. More generally $\text{Mat}_{n \times n}(\mathbb{R})$ (usually called $\mathfrak{gl}_n(\mathbb{R})$ in this context) is a Lie algebra with the bracket given by the matrix commutator: $[A, B] = AB - BA$. Notice that, by the previous example, $\mathfrak{so}(3)$ is a sub-Lie algebra of $\mathfrak{gl}_n(\mathbb{R})$.

Example 3.3.5. As you showed in HW 1 #4, for any manifold M , $\mathfrak{X}(M)$ is a Lie algebra, where the bracket is the Lie bracket.

By the last example, we know that for any Lie group G the collection of all vector fields $\mathfrak{X}(G)$ is a Lie algebra. However, this is not the Lie algebra we're after when we say *the* Lie algebra of G : first, because Lie

groups aren't special in this regard ($\mathfrak{X}(M)$ is a Lie algebra for any manifold), and second because $\mathfrak{X}(M)$ is a (non-trivial) $C^\infty(M)$ -module, and hence infinite-dimensional as a real vector space.

Building on Examples 1.5.3, 1.6.8, and 1.6.10, we can get a *finite-dimensional* sub-Lie algebra of $\mathfrak{X}(G)$ by focusing on the left-invariant vector fields:

Definition 3.3.6. A vector field X on G (by assumption not necessarily smooth) is *left-invariant* if, for all $g \in G$, $dL_g X = X$. More precisely, this means that for each $h \in G$,

$$(dL_g)_h(X(h)) = X(gh).$$

Theorem 3.3.7. Suppose G is a Lie group and \mathfrak{g} is the set of left-invariant vector fields on G . Then

- (i) \mathfrak{g} is a real vector space and the evaluation map $\text{ev}_e : \mathfrak{g} \rightarrow T_e G$ given by $\text{ev}_e(X) = X(e)$ is a vector space isomorphism. In particular, this implies that $\dim \mathfrak{g} = \dim T_e G = \dim G$.
- (ii) Elements of \mathfrak{g} are smooth vector fields.
- (iii) For $X, Y \in \mathfrak{g}$, $[X, Y] \in \mathfrak{g}$, where $[\cdot, \cdot]$ is the usual Lie bracket of vector fields on manifolds.
- (iv) $(\mathfrak{g}, [\cdot, \cdot])$ is a Lie algebra.

Proof. (i) The map ev_e is certainly linear. It's surjective since, if $v \in T_e G$, we can define a vector field X on G by $X(g) := (dL_g)_e v$. Then X is left-invariant (and hence in \mathfrak{g}) since

$$(dL_g)_h(X(h)) = (dL_g)_h((dL_h)_e v) = (d(L_g \circ L_h))_e(v) = (dL_{gh})_e(v) = X(gh).$$

But then $\text{ev}_e(X) = v$, so ev_e is surjective.

On the other hand, ev_e is also injective: if $\text{ev}_e(X) = \text{ev}_e(Y)$ for $X, Y \in \mathfrak{g}$, then for all $g \in G$ we have

$$X(g) = (dL_g)_e(X(e)) = (dL_g)_e(\text{ev}_e(X)) = (dL_g)_e(\text{ev}_e(Y)) = (dL_g)_e(Y(e)) = Y(g)$$

by left-invariance of X and Y , so we conclude that $\text{ev}_e(X) = \text{ev}_e(Y)$ implies $X = Y$.

(ii) is straightforward but annoying, so we skip it.

(iii) is a consequence of 3.3.8 stated below, and then (iv) is immediate given that the Lie bracket of vector fields is alternating and satisfies the Jacobi identity (which you proved in HW 1 #4). \square

Lemma 3.3.8. Suppose $f : M \rightarrow N$ is smooth and $X_M, Y_M \in \mathfrak{X}(M)$, $X_N, Y_N \in \mathfrak{X}(N)$ so that $df_p X_M(p) = X_N(f(p))$ and $df_p Y_M(p) = Y_N(f(p))$ for all $p \in M$. Then

$$df_p([X_M, Y_M](p)) = [X_N, Y_N](f(p)).$$

In the proof of Theorem 3.3.7 (iii), we apply this lemma with $M = N = G$ and $f = L_g$.

Proof of Lemma 3.3.8. Using the notation $Z_q = Z(q)$ to indicate the tangent vector determined by the vector field Z at a point q , it suffices to show that $(df_p[X_M, Y_M]_p)(\psi) = [X_N, Y_N]_{f(p)}(\psi)$ for any $\psi \in C^\infty(N)$.

This is a computation:

$$\begin{aligned}
(df_p[X_M, Y_M]_p)(\psi) &= [X_M, Y_M]_p(\psi \circ f) = (X_M)_p((Y_M)_p(\psi \circ f)) - (Y_M)_p((X_M)_p(\psi \circ f)) \\
&= (X_M)_p((df_p(Y_M)_p)(\psi)) - (Y_M)_p((df_p(X_M)_p)(\psi)) \\
&= (X_M)_p((Y_N)_{f(p)}(\psi)) - (Y_M)_p((X_N)_{f(p)}(\psi)) \\
&= (X_M)_p(Y_N(\psi) \circ f) - (Y_M)_p(X_N(\psi) \circ f) \\
&= (df_p X_M)(Y_N(\psi)) - (df_p Y_M)(X_N(\psi)) \\
&= (X_N)_{f(p)}(Y_N(\psi)) - (Y_N)_{f(p)}(X_N(\psi)) \\
&= [X_N, Y_N]_{f(p)}(\psi).
\end{aligned}$$

□

Theorem 3.3.7 tells us that any Lie group G has an associated finite-dimensional Lie algebra \mathfrak{g} of left-invariant vector fields; this is usually called *the Lie algebra of G* , typically denoted with a lowercase fraktur version of the name of the group, but sometimes also $\text{Lie}(G)$.

Notice, in particular, that we can also identify \mathfrak{g} with $T_e G$ using the evaluation map, so we could also interpret the Lie bracket as being an operation on this tangent space.

Example 3.3.9. Recall the group $\text{Aff}(\mathbb{R})$ of invertible affine transformations of \mathbb{R} from [Example 3.2.3](#), where we showed that we can represent $\text{Aff}(\mathbb{R})$ as the subgroup $\left\{ \begin{bmatrix} a & b \\ 0 & 1 \end{bmatrix} : a \neq 0 \right\} \subset \text{GL}_2(\mathbb{R})$.

In this representation, $\mathfrak{aff}(\mathbb{R}) \cong T_I \text{Aff}(\mathbb{R})$ has vector space basis $A := \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, $B := \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$; by abuse of notation we can either think of A and B as these specific matrices, or as the left-invariant vector fields on $\text{Aff}(\mathbb{R})$ that they determine.

Exercise 3.3.10. There is an obvious global coordinate chart $\phi: P \rightarrow \text{Aff}(\mathbb{R})$, where $P = \{(u, v) \in \mathbb{R}^2 : u \neq 0\}$ is the plane minus the y -axis, given by $\phi(u, v) = \begin{bmatrix} u & v \\ 0 & 1 \end{bmatrix}$, and associated local coordinate basis $\frac{\partial}{\partial u}, \frac{\partial}{\partial v}$ for each tangent space. The left-invariant vector fields A, B can be written in this basis as

$$A \begin{bmatrix} u & v \\ 0 & 1 \end{bmatrix} = u \frac{\partial}{\partial u} \quad \text{and} \quad B \begin{bmatrix} u & v \\ 0 & 1 \end{bmatrix} = u \frac{\partial}{\partial v}. \quad (3.2)$$

Continuing the example, we will shortly see ([Proposition 3.4.2](#)) that the Lie bracket in the Lie algebra of any matrix group (where we interpret the Lie algebra as the tangent space to the identity, which is just some space of matrices) can be computed as the matrix commutator. Therefore,

$$[A, B] = AB - BA = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} - \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} - \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = B.$$

Anti-commutativity implies $[A, A] = 0$ and $[B, B] = 0$, so $[A, B] = B$ completely determines the Lie algebra structure on $\mathfrak{aff}(\mathbb{R})$. In fact, this turns out to be the *only* non-abelian 2-dimensional Lie algebra.¹

¹That is, the only other 2-dimensional Lie algebra has basis X, Y and trivial bracket $[X, Y] = 0$; for example, this is the Lie algebra of the torus $U(1)^2$, which can be represented by the group of diagonal unitary 2×2 matrices $\left\{ \begin{bmatrix} e^{i\theta_1} & 0 \\ 0 & e^{i\theta_2} \end{bmatrix} \right\} \subset U(2)$.

This Lie algebra is *solvable*, since the derived series terminates:

$$\mathfrak{aff}(\mathbb{R})^{(1)} := [\mathfrak{aff}(\mathbb{R}), \mathfrak{aff}(\mathbb{R})] = \langle B \rangle \quad \text{and} \quad \mathfrak{aff}(\mathbb{R})^{(2)} := [\mathfrak{aff}(\mathbb{R})^{(1)}, \mathfrak{aff}(\mathbb{R})^{(1)}] = [\langle B \rangle, \langle B \rangle] = \{0\}.$$

It is not *nilpotent* since the lower central series continues forever:

$$\mathfrak{aff}(\mathbb{R})_{(1)} := [\mathfrak{aff}(\mathbb{R}), \mathfrak{aff}(\mathbb{R})] = \langle B \rangle, \quad \mathfrak{aff}(\mathbb{R})_{(2)} := [\mathfrak{aff}(\mathbb{R}), \mathfrak{aff}(\mathbb{R})_{(1)}] = [\mathfrak{aff}(\mathbb{R}), \langle B \rangle] = \langle B \rangle, \quad \text{etc.}$$

If α and β are the dual left-invariant 1-forms on $\text{Aff}(\mathbb{R})$, then using [Theorem 2.7.5](#) (Cartan's Magic Formula),

$$\begin{aligned} d\alpha(A, B) &= (\iota_A d\alpha)(B) = (\mathcal{L}_A \alpha)(B) - (d(\iota_A \alpha))(B) = \mathcal{L}_A(\alpha(B)) - \alpha(\mathcal{L}_A B) - (d(1))(B) \\ &= \mathcal{L}_A(0) - \alpha([A, B]) - 0 = -\alpha(B) = 0 \end{aligned}$$

and

$$\begin{aligned} d\beta(A, B) &= (\iota_A d\beta)(B) = (\mathcal{L}_A \beta)(B) - (d(\iota_A \beta))(B) = \mathcal{L}_A(\beta(B)) - \beta(\mathcal{L}_A B) - (d(0))(B) \\ &= \mathcal{L}_A(1) - \beta([A, B]) - 0 = -\beta(B) = 1, \end{aligned}$$

so $d\alpha = 0$ and $d\beta = -\alpha \wedge \beta$.

Just by looking at the coordinate expressions (3.2) for A and B , we could pretty easily have written down the formulas

$$\alpha \begin{bmatrix} u & v \\ 0 & 1 \end{bmatrix} = \frac{1}{u} du \quad \text{and} \quad \beta \begin{bmatrix} u & v \\ 0 & 1 \end{bmatrix} = \frac{1}{u} dv,$$

from which we could have directly computed

$$\begin{aligned} d\alpha &= d\left(\frac{1}{u} du\right) = 0 \\ d\beta &= d\left(\frac{1}{u} dv\right) = -\frac{1}{u^2} du \wedge dv = -\left(\frac{1}{u} du\right) \wedge \left(\frac{1}{u} dv\right) = -\alpha \wedge \beta. \end{aligned}$$

From this perspective, it's clear that $\alpha = d(\log |u|)$.²

3.4 Matrix Groups

Definition 3.4.1. A *matrix group* is a closed subgroup of $\text{GL}_n(\mathbb{R})$ for some n .

You might worry that this excludes groups consisting of complex or quaternionic matrices, but all such groups can be represented as subgroups of some real general linear group. Specifically, define the map $\rho_1: \mathbb{C} \rightarrow \text{Mat}_{2 \times 2}(\mathbb{R})$ by

$$\rho_1(x + yi) = \begin{bmatrix} x & -y \\ y & x \end{bmatrix}.$$

²Because the base-10 logarithm is pretty useless and I'm not a computer scientist, I use \log to denote the natural logarithm (a.k.a. \ln).

This is the right definition to turn multiplication of complex numbers into linear transformations of \mathbb{R} . After all, if $z = x + iy$, then for $w = u + iv$ we have

$$zw = (x + yi)(u + vi) = xu - yv + i(xv + yu)$$

and

$$\rho_1(z) \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} x & -y \\ y & x \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} xu - yv \\ xv + yu \end{bmatrix}.$$

In turn, we can get a map $\rho_n: \text{Mat}_{n \times n}(\mathbb{C}) \rightarrow \text{Mat}_{2n \times 2n}(\mathbb{R})$ by just applying ρ_1 to each entry, resulting in a real matrix built out of 2×2 blocks of the form $\begin{bmatrix} x & -y \\ y & x \end{bmatrix}$. For example

$$\rho_3 \left(\begin{bmatrix} x_{11} + y_{11}i & x_{12} + y_{12}i & x_{13} + y_{13}i \\ x_{21} + y_{21}i & x_{22} + y_{22}i & x_{23} + y_{23}i \\ x_{31} + y_{31}i & x_{32} + y_{32}i & x_{33} + y_{33}i \end{bmatrix} \right) = \begin{bmatrix} x_{11} & -y_{11} & x_{12} & -y_{12} & x_{13} & -y_{13} \\ y_{11} & x_{11} & y_{12} & x_{12} & y_{13} & x_{13} \\ x_{21} & -y_{21} & x_{22} & -y_{22} & x_{23} & -y_{23} \\ y_{21} & x_{21} & y_{22} & x_{22} & y_{23} & x_{23} \\ x_{31} & -y_{31} & x_{32} & -y_{32} & x_{33} & -y_{33} \\ y_{31} & x_{31} & y_{32} & x_{32} & y_{33} & x_{33} \end{bmatrix}.$$

It's not hard to show that ρ_n preserves invertibility, so $\rho_n(\text{GL}_n(\mathbb{C})) \subset \text{GL}_{2n}(\mathbb{R})$, showing that $\text{GL}_n(\mathbb{C})$ and all of its closed subgroups are matrix groups.

Similarly, we can define $\psi_1: \mathbb{H} \rightarrow \text{Mat}_{2 \times 2}(\mathbb{C})$ by

$$\psi_1(z + wj) = \begin{bmatrix} z & w \\ -\bar{w} & \bar{z} \end{bmatrix}$$

or, written out more completely,

$$\psi_1(a + bi + cj + dk) = \begin{bmatrix} a + bi & c + di \\ -c + di & a - bi \end{bmatrix}.$$

And then we can extend to a map $\psi_n: \text{Mat}_{n \times n}(\mathbb{H}) \rightarrow \text{Mat}_{2n \times 2n}(\mathbb{C})$ as you'd expect. For example,

$$\begin{aligned} \psi_2 \left(\begin{bmatrix} a_{11} + b_{11}i + c_{11}j + d_{11}k & a_{12} + b_{12}i + c_{12}j + d_{12}k \\ a_{21} + b_{21}i + c_{21}j + d_{21}k & a_{22} + b_{22}i + c_{22}j + d_{22}k \end{bmatrix} \right) \\ = \begin{bmatrix} a_{11} + ib_{11} & c_{11} + id_{11} & a_{12} + ib_{12} & c_{12} + id_{12} \\ -c_{11} + id_{11} & a_{11} - ib_{11} & -c_{12} + id_{12} & a_{12} - ib_{12} \\ a_{21} + ib_{21} & c_{21} + id_{21} & a_{22} + ib_{22} & c_{22} + id_{22} \\ -c_{21} + id_{21} & a_{21} - ib_{21} & -c_{22} + id_{22} & a_{22} - ib_{22} \end{bmatrix} \end{aligned}$$

Again, $\psi_n(\text{GL}_n(\mathbb{H})) \subset \text{GL}_{2n}(\mathbb{C})$, and of course by composing ψ_n with ρ_{2n} we can represent any quaternionic matrix group as a subgroup of $\text{GL}_{4n}(\mathbb{R})$.

The nice thing about matrix groups is that the bracket operations on their Lie algebras are particularly simple.

Proposition 3.4.2. *Let $G \leq \text{GL}_n(\mathbb{R})$ be a matrix group, so that $T_I G \subset T_I \text{GL}_n(\mathbb{R}) \cong \text{Mat}_{n \times n}(\mathbb{R})$. Then the bracket operation on \mathfrak{g} , interpreted as an operation on $T_I G$, is just the matrix commutator: for $X, Y \in T_I G \cong \mathfrak{g}$, the bracket $[X, Y] = XY - YX$ where XY just means the product of matrices.*

Proof. We prove this for $G = \mathrm{GL}_n(\mathbb{R})$ and then it follows for subgroups.

We're going to prove this by writing out the matrix multiplication in terms of coordinates; i.e., if $A, B \in \mathrm{Mat}_{n \times n}(\mathbb{R})$, then

$$(AB)_{ij} = \sum_{k=1}^n A_{ik}B_{kj}.$$

But now, if we're thinking of our matrices as tangent vectors at the identity, then they're really differential operators, so we need a way of interpreting these coordinates in that language.

To that end, by analogy with what we did in [Section 2.4.1](#), define $x_{ij} : \mathrm{Mat}_{n \times n}(\mathbb{R}) \rightarrow \mathbb{R}$ by $x_{ij}(A) = A_{ij}$, the (i, j) entry of A . Then if $V \in T_I \mathrm{GL}_n(\mathbb{R}) \cong \mathrm{Mat}_{n \times n}(\mathbb{R})$, we can write $V = \alpha'(0)$ where $\alpha(t) = I + tV$, which stays in $\mathrm{GL}_n(\mathbb{R})$ for small enough t . Therefore, using [Definition 1.2.1](#)

$$V(x_{ij}) = (x_{ij} \circ \alpha)'(0) = \frac{d}{dt} \Big|_{t=0} x_{ij}(\alpha(t)) = \frac{d}{dt} \Big|_{t=0} \alpha(t)_{ij} = \alpha'(0)_{ij} = V_{ij}. \quad (3.3)$$

So this gives a way of translating back and forth between elements of $T_I \mathrm{GL}_n(\mathbb{R})$ as differential operators and as arrays of numbers.

Next, recall the evaluation map $\mathrm{ev}_I : \mathfrak{gl}_n(\mathbb{R}) \rightarrow T_I \mathrm{GL}_n(\mathbb{R})$ given by $\mathrm{ev}_I(X) = X(I)$, the tangent vector at I determined by the left-invariant vector field X . In what follows, to avoid a surplus of parentheses, we will sometimes write $X(I)$ as X_I .

We saw in the proof of [Theorem 3.3.7](#) that ev_I is a vector space isomorphism, so all that remains to do is to show that it is a Lie algebra homomorphism; that is, for any left-invariant vector fields X, Y on $\mathrm{GL}_n(\mathbb{R})$ (i.e., elements of $\mathfrak{gl}_n(\mathbb{R})$), we need to show that

$$\mathrm{ev}_I([X, Y]) = \mathrm{ev}_I(X) \mathrm{ev}_I(Y) - \mathrm{ev}_I(Y) \mathrm{ev}_I(X) = X(I)Y(I) - Y(I)X(I),$$

where on the right hand side we're just multiplying matrices.

Notice, first of all, that for any $g, h \in \mathrm{GL}_n(\mathbb{R})$,

$$(x_{ij} \circ L_g)(h) = x_{ij}(L_g(h)) = x_{ij}(gh) = (gh)_{ij} = \sum_{k=1}^n g_{ik}h_{kj} = \sum_{k=1}^n x_{ik}(g)x_{kj}(h);$$

in other words,

$$x_{ij} \circ L_g = \sum_{k=1}^n x_{ik}(g)x_{kj}. \quad (3.4)$$

We're going to look now at the function $Y(x_{ij})$, which we can specify by saying how it evaluates at each $g \in \mathrm{GL}_n(\mathbb{R})$. By left-invariance of Y together with (3.3) and (3.4),

$$(Y(x_{ij})) (g) = Y_g(x_{ij}) = ((dL_g)_I Y_I)(x_{ij}) = Y_I(x_{ij} \circ L_g) = \sum_{k=1}^n x_{ik}(g)Y_I(x_{kj}) = \sum_{k=1}^n x_{ik}(g)Y(I)_{kj}.$$

This (and the corresponding statement for X) gives us the necessary tool to determine the (i, j) component of $\mathrm{ev}_I([X, Y])$:

$$\begin{aligned} (\mathrm{ev}_I([X, Y]))_{ij} &= [X, Y]_I(x_{ij}) = X_I(Y(x_{ij})) - Y_I(X(x_{ij})) \\ &= X_I \left(\sum_{k=1}^n x_{ik}Y(I)_{kj} \right) - Y_I \left(\sum_{k=1}^n x_{ik}X(I)_{kj} \right) = \sum_{k=1}^n (X_I(x_{ik}Y(I)_{kj}) - Y_I(x_{ik}X(I)_{kj})) \end{aligned}$$

by linearity of the operators X_I and Y_I . Of course, $Y(I)_{kj}$ and $X(I)_{kj}$ are constant, so the Leibniz rule gives us

$$\sum_{k=1}^n (X_I(x_{ik})Y(I)_{kj} - Y_I(x_{ik})X(I)_{kj}) = \sum_{k=1}^n (X(I)_{ik}Y(I)_{kj} - Y(I)_{ik}X(I)_{kj}) = (X(I)Y(I) - Y(I)X(I))_{ij},$$

where, as a reminder, I'm freely switching between the notations $X(I)$ and X_I for the tangent vector at I determined by the vector field X . So this shows that the left and right sides of the equation in the statement of the theorem agree componentwise. \square

This now justifies the computational simplifications we did in looking at the Lie algebras of $U(d)$ (Example 1.6.8), $SO(d)$ (Example 1.6.10), and $Aff(\mathbb{R})$ (Example 3.3.9).

Proposition 3.4.3. $\mathfrak{so}(3) \cong \mathfrak{su}(2) \cong \mathfrak{sp}(1)$.

Proof. In each case this is the Lie algebra of a matrix group, so Proposition 3.4.2 guarantees it is isomorphic to the tangent space at the identity with the bracket given by the matrix commutator. This is the interpretation we will use to prove the proposition.

For example, $T_I SO(3)$ consists of the 3×3 skew-symmetric matrices, which we saw in Example 1.6.10 is isomorphic to (\mathbb{R}^3, \times) .

We've seen in Example 1.5.3 that $T_I U(d)$ consists of the skew-Hermitian $d \times d$ matrices; I claim without proof that the determinant-1 condition translates to a trace-0 condition, so $T_I SU(d)$ is the space of traceless skew-Hermitian matrices.

Finally, thinking of $Sp(1) \cong S^3$ as the 1×1 quaternionic matrices which preserve the symplectic inner product (i.e., the quaternionic analog of $O(d)$ or $U(d)$), the same sort of argument that gave the tangent spaces at the identity of $O(d)$ and $U(d)$ will imply that $T_I Sp(1)$ consists of quaternionic skew-Hermitian 1×1 matrices; in other words, the single entry must be a purely imaginary quaternion $ai + bj + ck$.

We can fairly easily write down bases for $\mathfrak{so}(3)$, $\mathfrak{su}(2)$, and $\mathfrak{sp}(1)$:

$$\begin{aligned}\mathfrak{so}(3) &= \text{span} \left\{ \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \right\} \\ \mathfrak{su}(2) &= \text{span} \left\{ \frac{1}{2} \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}, \frac{1}{2} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \frac{1}{2} \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix} \right\} \\ \mathfrak{sp}(1) &= \text{span} \left\{ \frac{i}{2}, \frac{j}{2}, \frac{k}{2} \right\}.\end{aligned}$$

If we use $\{A_{1,g}, A_{2,g}, A_{3,g}\}$ to denote the basis for $g \in \{\mathfrak{so}(3), \mathfrak{su}(2), \mathfrak{sp}(1)\}$, then it is easy to check that

$$[A_{1,g}, A_{2,g}] = A_{3,g}, \quad [A_{2,g}, A_{3,g}] = A_{1,g}, \quad [A_{3,g}, A_{1,g}] = A_{2,g}.$$

For example,

$$[A_{1,\mathfrak{so}(3)}, A_{2,\mathfrak{so}(3)}] = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix} - \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = A_{3,\mathfrak{so}(3)}$$

and

$$[A_{3,\mathfrak{sp}(1)}, A_{1,\mathfrak{sp}(1)}] = \frac{k}{2} \frac{i}{2} - \frac{i}{2} \frac{k}{2} = \frac{j}{4} + \frac{j}{4} = \frac{j}{2} = A_{2,\mathfrak{sp}(1)}.$$

Therefore, for $\{\mathfrak{g}, \mathfrak{g}'\} \subset \{\mathfrak{so}(3), \mathfrak{su}(2), \mathfrak{sp}(1)\}$, linearly extending the identifications $A_{i,\mathfrak{g}} \mapsto A_{i,\mathfrak{g}'}$ defines a Lie algebra isomorphism. \square

[Proposition 3.4.3](#) is an example of what is sometimes called an *accidental isomorphism*, which in this case boil down to the fact that there just aren't that many 3-dimensional Lie algebras. In general $\mathfrak{su}(n+1)$, $\mathfrak{so}(2n+1)$, and $\mathfrak{sp}(n)$ belong to three different infinite families of Lie algebras: they are A_n , B_n , and C_n , respectively. In this language, [Proposition 3.4.3](#) says that $A_1 \cong B_1 \cong C_1$.

Bibliography

- [1] Simon K. Donaldson. An application of gauge theory to four-dimensional topology. *Journal of Differential Geometry*, 18(2):279–315, 1983.
- [2] Michael Hartley Freedman. The topology of four-dimensional manifolds. *Journal of Differential Geometry*, 17(3):357–453, 1982.
- [3] Robert E. Gompf. Three exotic \mathbf{R}^4 's and other anomalies. *Journal of Differential Geometry*, 18(2):317–328, 1983.
- [4] Victor Guillemin and Alan Pollack. *Differential Topology*. Number 370 in AMS Chelsea Publishing. American Mathematical Society, Providence, RI, USA, 2010.
- [5] Brian C. Hall. *Lie Groups, Lie Algebras, and Representations: An Elementary Introduction*. Number 222 in Graduate Texts in Mathematics. Springer, Cham, 2nd ed. 2015 edition, 2015.
- [6] Michel A. Kervaire and John W. Milnor. Groups of homotopy spheres: I. *The Annals of Mathematics*, 77(3):504, May 1963.
- [7] John Milnor. On manifolds homeomorphic to the 7-sphere. *The Annals of Mathematics*, 64(2):399–405, September 1956.
- [8] Clifford Henry Taubes. Gauge theory on asymptotically periodic 4-manifolds. *Journal of Differential Geometry*, 25(3):363–430, 1987.
- [9] Hassler Whitney. The self-intersections of a smooth n -manifold in $2n$ -space. *The Annals of Mathematics*, 45(2):220–246, April 1944.
- [10] Hassler Whitney. The singularities of a smooth n -manifold in $(2n-1)$ -space. *The Annals of Mathematics*, 45(2):247–293, April 1944.