Introduction to Differential Manifolds

Clayton Shonkwiler

Department of Mathematics, Colorado State University, Fort Collins, CO, USA

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 $\phi_\alpha : U_\alpha \to M$ 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 $\phi_\beta^{-1}(W) \subseteq \mathbb{R}^n$.

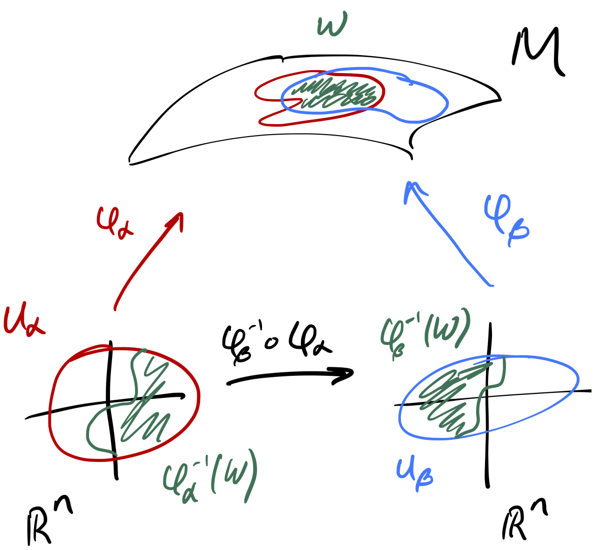

Figure 1.1: Transition maps.

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 $(\mathbb{R}^n, \mathrm{id})$ 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."

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

[2] In 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 [6]; it turns out there are exactly 28 distinct smooth structures on S^7 [5]) 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 $\mathbb{R}^4$ [7]). 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 $M \times M \times \cdots \times M$.

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

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

and

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

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 $(x_1, \ldots, x_n, 0) \in \mathbb{R}^{n+1}$ to the north pole $(0, \ldots, 0, 1) \in S^n \subset \mathbb{R}^{n+1}$; see Figure 1.2.

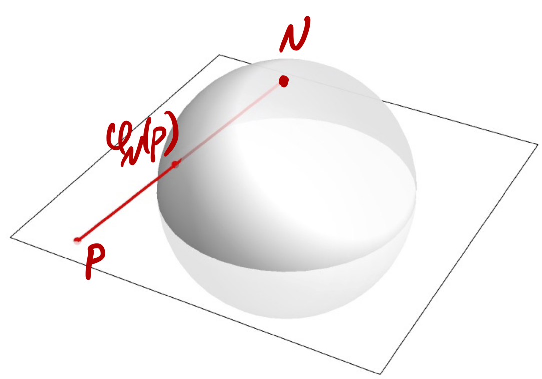

Figure 1.2: Inverse stereographic projection.

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

For (ii), observe that

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

so

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

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

The inverse of ϕ_N is stereographic projection

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

(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 $f : M \to N$ is *differentiable* at $p \in M$ if, given a coordinate chart $\psi : V \subseteq \mathbb{R}^n \to N$ containing $f(p)$, there exists a coordinate chart $\phi : U \subseteq \mathbb{R}^m \to M$ containing p so that $f(\phi(U)) \subseteq \psi(V)$ and

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

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 $\alpha : S^n \to S^n$ 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 $\alpha(\phi_S(\mathbb{R}^n)) = S^n \backslash \{N\} = \phi_N(\mathbb{R}^n)$. 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 $\mathbb{R}^n \to \mathbb{R}^n$).

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.

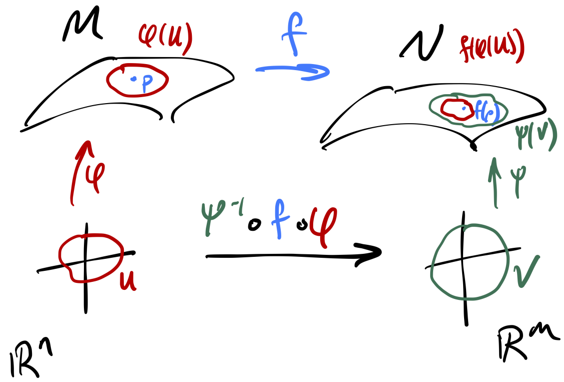

Figure 1.3: Locally converting a map between manifolds to a map between open sets in Euclidean spaces, where we know what differentiability means.

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 [8, 9] 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 $\alpha : (-\epsilon, \epsilon) \to \mathbb{R}^n$ 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 $x_i : (-\epsilon, \epsilon) \to \mathbb{R}$ are themselves smooth, then

$$\alpha'(0) = (x_1'(0), \dots, x_n'(0)) = \vec{v}.$$

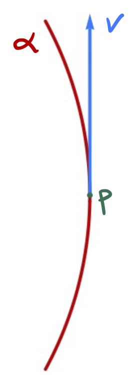

Figure 1.4: A curve through p with velocity $\vec{v}$ at p.

(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 $f : U \to \mathbb{R}$ 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

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

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

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

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

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

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 $\alpha\colon(-\epsilon,\epsilon)\to M$ 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 $x_1,\dots,x_n\colon(-\epsilon,\epsilon)\to\mathbb{R}$. Then

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

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 $\alpha'(0)\in T_pM$ in local coordinates as

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

Since the $x_i'(0)$ are just scalars, what we're doing here is writing $\alpha'(0)$ in terms of the *local coordinate basis* $\left\{\left(\frac{\partial}{\partial x_1}\right)_{\vec{0}},\dots,\left(\frac{\partial}{\partial x_n}\right)_{\vec{0}}\right\}$ for T_pM 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 $\left\{\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_n}\right\}$ 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_pM and T_qM 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

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

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

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

Notice that there are natural projections $\pi : TM \to M$ and $\widetilde{\pi} : T^*M \to M$ 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 $\mathrm{SO}(3) \cong \mathbb{RP}^3$, 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 $\pi_1(\mathbb{RP}^3) \cong \mathbb{Z}/2\mathbb{Z}$.)

[3] Meaning 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 $S^1 \times \cdots \times S^1 = (S^1)^n$.

1.3 The Differential

In multivariable calculus, any differentiable map $f : \mathbb{R}^m \to \mathbb{R}^n$ has a corresponding differential df.[4] While in a multivariable calculus setting we often think of df as a map $\mathbb{R}^m \to \mathbb{R}^n$ 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_pM 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 $f : \mathbb{R}^m \to \mathbb{R}^n$ and a point p in the domain of f, the differential of f at p is a map $df_p : T_p\mathbb{R}^m \to T_{f(p)}\mathbb{R}^n$, 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 $f : M \to N$ is smooth. For each $p \in M$, define the *differential* of f at p, denoted $df_p : T_pM \to T_{f(p)}N$, as follows: For any $v \in T_pM$, choose a smooth curve $\alpha : (-\epsilon, \epsilon) \to M$ 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 $f : M \to \mathbb{R}$ 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:

[4] You 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, $p \in M$, $v \in T_pM$, and that $f : M \to \mathbb{R}$ is differentiable in a neighborhood of p. Then*

$$vf = df_p v.$$

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

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

But this is exactly the definition of $df_p v$. □

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

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

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

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

See Figure 1.5.

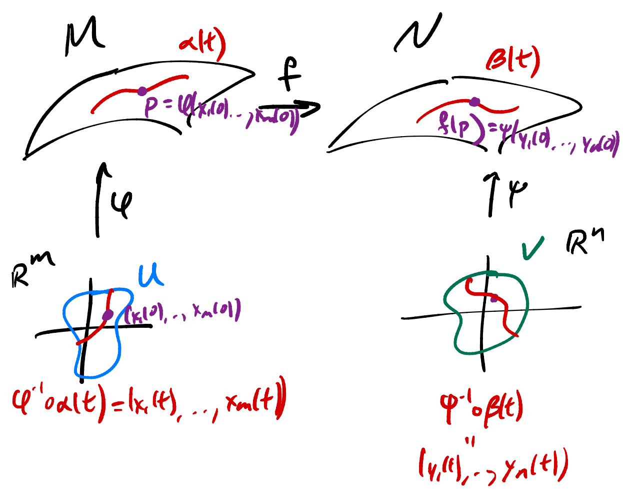

Figure 1.5: The differential in local coordinates

Of course,

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

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

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

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

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

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

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

where $\left[\frac{\partial y_i}{\partial x_j}\right]_{i,j}$ 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 $f : M \to \mathbb{R}$. Then, at a point $p \in M$, the differential $df_p : T_pM \to T_{f(p)}\mathbb{R}$. 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 $T_pM \to \mathbb{R}$.

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_pM. Or, in fancier language, df_p is an element of the dual space $(T_pM)^*$. 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 $f : M \to N$, we can talk about critical points and critical values, and doing so is often quite important.

Definition 1.3.5. Let $f : M \to N$ be smooth. A point $p \in M$ is a *critical point* of f if $df_p : T_pM \to T_{f(p)}N$ 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

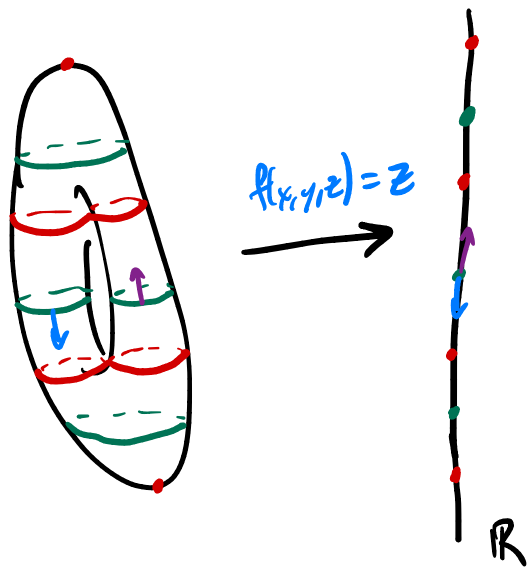

Figure 1.6: The height function on a torus.

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.

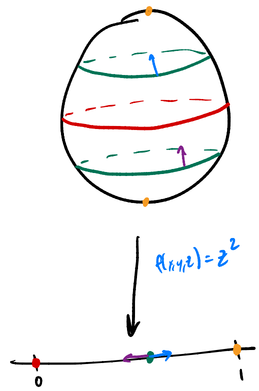

Figure 1.7: A smooth function on the sphere which does not have isolated critical points.

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 $m \geq n$, $f : M^m \to N^n$ 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 $f : \mathbb{R}^n \to \mathbb{R}$ by $f(x_1, \ldots, x_n) = x_1^2 + \cdots + x_n^2$. Then

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

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 $\mathrm{Mat}_{d\times N}(\mathbb{C})$ be the space of $d\times N$ complex matrices. This is trivially a $2dN$-dimensional manifold, since $\mathrm{Mat}_{d\times N}(\mathbb{C}) \cong \mathbb{C}^{dN}$, 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 $\Phi\colon \mathrm{Mat}_{d\times N}(\mathbb{C}) \to \mathcal{H}(d)$ by

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

(In frame theory, we would think of $A \in \mathrm{Mat}_{d\times N}(\mathbb{C})$ 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 $\mathrm{St}_d(\mathbb{C}^N)$ 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 $\mathrm{U}(d)$! So we've proved that $\mathrm{U}(d)$ 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 $\mathrm{O}(d)$ 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 $f\colon M \to N$ is smooth. Then f is an *immersion* if df_p is injective for all $p \in M$ (note that this implies $\dim(M) \le \dim(N)$).

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 $f\colon U \to V$ 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 $\alpha\colon \mathbb{R} \to \mathbb{R}^2$ 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 $\beta\colon \mathbb{R} \to \mathbb{R}^2$ 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

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

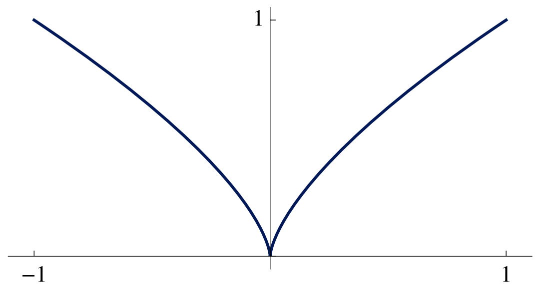

Figure 1.8: The image of the map α from Example 1.4.2. The fact that it has a cusp suggests that it will fail to be an immersion, despite being injective and smooth.

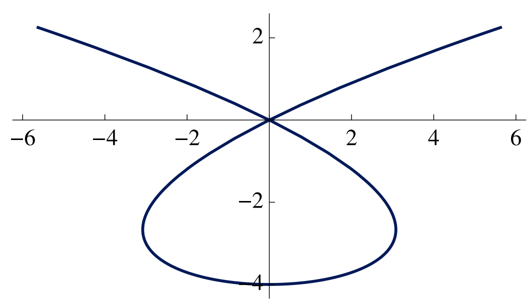

Figure 1.9: The image of the map β from Example 1.4.3. This map is an immersion, but not an embedding.

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 $\phi\colon U \subset \mathbb{R}^n \to M$ 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, $\phi\colon U \to \phi(U)$ is smooth because $\phi^{-1} \circ \phi \circ \mathrm{id} = \mathrm{id}$ is smooth everywhere on U. Similarly, $\phi^{-1}\colon \phi(U) \to U$ is smooth because $\mathrm{id} \circ \phi^{-1} \circ \phi = \mathrm{id}$ is smooth everywhere on U.)

If $f\colon M \to N$ is a local diffeomorphism at $p \in M$, then $df_p\colon T_pM \to T_{f(p)}N$ 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 $f\colon M \to N$ is smooth. If $p \in M$ and $df_p\colon T_pM \to T_{f(p)}N$ 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 $\psi^{-1} \circ f \circ \phi\colon \phi(U) \to \psi(V)$, 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,

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

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 $F\colon U \subseteq \mathbb{R}^n \to \mathbb{R}^n$ is given by $(x_1, \dots, x_n) \mapsto (y_1, \dots, y_n)$, then $dF_p = \left[\left.\frac{\partial y_i}{\partial x_j}\right|_p \right]_{i,j}$ 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 $f : M \to N$ 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

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

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.

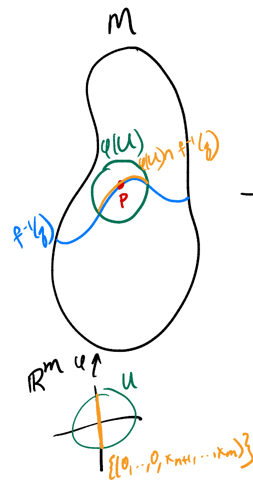

Figure 1.10: A sketch of the key neighborhoods and maps in the proof of Theorem 1.3.9.

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. $\square$

[5]Meaning 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 $X : M \to TM$. 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 $\pi : TM \to M$ which maps $v \in T_pM$ 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 $\sigma : B \to E$ so that $\pi \circ \sigma = \mathrm{id}_B$, 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 $X : M \to TM$ is a smooth map satisfying $\pi \circ X = \mathrm{id}_M$. 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 $\phi: U \subseteq \mathbb{R}^n \to M$ 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 $a_i: U \to \mathbb{R}$ 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 $C^\infty(M) \to C^\infty(M)$ 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 $\mathrm{U}(d)$ 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 $\mathrm{U}(d)$.

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

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

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

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 $T_I \mathrm{U}(d)$ are precisely the skew-Hermitian matrices.

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

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

$$(dL_U)_I : T_I \mathrm{U}(d) \to T_U \mathrm{U}(d)$$

and so, for any $\Delta \in T_I \mathrm{U}(d)$, we get a vector field $X_\Delta \in \mathfrak{X}(\mathrm{U}(d))$ on $\mathrm{U}(d)$ 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 $(dL_U)_I : T_I \,\mathrm{U}(d) \to T_U \,\mathrm{U}(d)$ 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 $\mathrm{U}(d) \to T_I \,\mathrm{U}(d)$ 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 $\exp : T_I \,\mathrm{U}(d) \to \mathrm{U}(d)$.

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

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

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 [4, Chapter 5].)

Claim. $\exp\colon T_I\,\mathrm{U}(d) \to \mathrm{U}(d)$ 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

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

(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 $H = V\operatorname{diag}(i\theta_1, \dots, i\theta_d)V^*$ 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. □

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

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

Proof. By definition,

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

□

To recap, we have a surjective map $\exp\colon T_I\,\mathrm{U}(d) \to \mathrm{U}(d)$. 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 $T_I\,\mathrm{U}(d)$ consisting of skew-Hermitian matrices with spectral norm $< \pi$ bijectively onto the neighborhood of I in $\mathrm{U}(d)$ 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 $\mathrm{U}(d)$. 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.

[6]If 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.

---

[7] Pronunciation 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."

[8] This correpondence 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.

[9] Interpreting 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.

[10] Look 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 $\left\{\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_n}\right\}$, 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

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

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 $\left\{\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_n}\right\}$.

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

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

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

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

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

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

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

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

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 $\phi\colon (0, 2\pi) \times (-1, 1) \to S^2$ given by

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

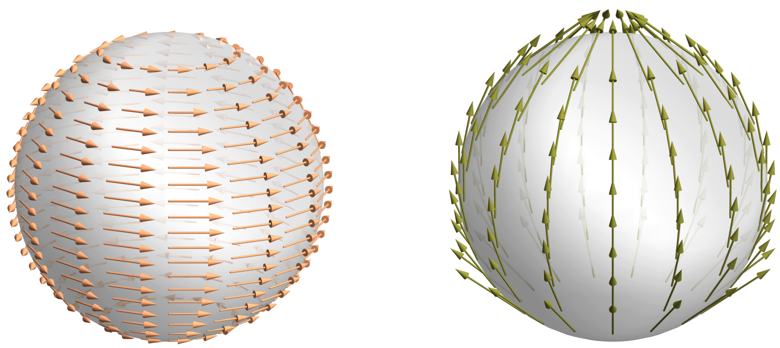

Figure 1.11: The vector fields $X = \frac{\partial}{\partial\theta}$ and $Y = \frac{\partial}{\partial z}$ on S^2.

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 $\mathrm{U}(d)$ are of the form $X(U) = U\Delta_X$, for $\Delta_X \in T_I\,\mathrm{U}(d)$, 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, it will turn out that the Lie bracket on left-invariant vector fields on a matrix group like $\mathrm{U}(d)$ 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 $U \in \mathrm{U}(d)$,

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

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 $T_I\,\mathrm{U}(d)$ turns the Lie bracket of left-invariant vector fields into the matrix commutator in $T_I\,\mathrm{U}(d)$. 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 $\mathrm{SO}(d)$, 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 $\mathrm{SO}(d)$ 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 $F : \mathbb{R}^3 \to \mathfrak{so}(3)$ given by

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

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

$$[\Delta_1, \Delta_2] = \Delta_1\Delta_2 - \Delta_2\Delta_1 = \begin{bmatrix} 0 & x_2y_1 - x_1y_2 & x_2z_1 - x_1z_2 \\ x_1y_2 - x_2y_1 & 0 & y_2z_1 - y_1z_2 \\ x_1z_2 - x_2z_1 & y_1z_2 - y_2z_1 & 0 \end{bmatrix}.$$

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 [4, 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.)

Bibliography

[1] Simon K. Donaldson. An application of gauge theory to four-dimensional topology. *Journal of Differential Geometry*, 18(2), 1983. DOI: 10.4310/jdg/1214437665 (↑ 3).

[2] Michael Hartley Freedman. The topology of four-dimensional manifolds. *Journal of Differential Geometry*, 17(3), 1982. DOI: 10.4310/jdg/1214437136 (↑ 3).

[3] Robert E. Gompf. Three exotic $\mathbf{R}^4$'s and other anomalies. *Journal of Differential Geometry*, 18(2), 1983. DOI: 10.4310/jdg/1214437666 (↑ 3).

[4] 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. DOI: 10.1007/978-3-319-13467-3 (↑ 21, 28).

[5] Michel A. Kervaire and John W. Milnor. Groups of homotopy spheres: I. *The Annals of Mathematics*, 77(3):504, 1963. DOI: 10.2307/1970128. JSTOR: 1970128 (↑ 3).

[6] John Milnor. On manifolds homeomorphic to the 7-sphere. *The Annals of Mathematics*, 64(2):399, 1956. DOI: 10.2307/1969983. JSTOR: 1969983 (↑ 3).

[7] Clifford Henry Taubes. Gauge theory on asymptotically periodic 4-manifolds. *Journal of Differential Geometry*, 25(3), 1987. DOI: 10.4310/jdg/1214440981 (↑ 3).

[8] Hassler Whitney. The self-intersections of a smooth n-manifold in $2n$-space. *The Annals of Mathematics*, 45(2):220, 1944. DOI: 10.2307/1969265. JSTOR: 1969265 (↑ 5).

[9] Hassler Whitney. The singularities of a smooth n-manifold in $(2n-1)$-space. *The Annals of Mathematics*, 45(2):247, 1944. DOI: 10.2307/1969266. JSTOR: 1969266 (↑ 5).